US 20240200102A1

(19) **United States**

(12) **Patent Application Publication**

Kim

(10) **Pub. No.: US 2024/0200102 A1**

(43) **Pub. Date: Jun. 20, 2024**

(54) **FUSION PROTEINS COMPRISING AN INTEIN POLYPEPTIDE AND METHODS OF USE THEREOF**

(71) Applicant: **Pairwise Plants Services, Inc.**, Durham, NC (US)

(72) Inventor: **Yongjoo Kim**, Chapel Hill, NC (US)

(21) Appl. No.: **18/541,481**

(22) Filed: **Dec. 15, 2023**

Related U.S. Application Data

(60) Provisional application No. 63/387,767, filed on Dec. 16, 2022.

Publication Classification

(51) **Int. Cl.**

C12N 15/90	(2006.01)
C07K 14/195	(2006.01)
C12N 9/12	(2006.01)
C12N 9/22	(2006.01)
C12N 15/113	(2006.01)

(52) **U.S. Cl.**

CPC ***C12N 15/90*** (2013.01); ***C07K 14/195*** (2013.01); ***C12N 9/1276*** (2013.01); ***C12N 9/22*** (2013.01); ***C12N 15/113*** (2013.01); *C07K 2319/92* (2013.01); *C12N 2310/20* (2017.05); *C12N 2800/80* (2013.01)

(57) **ABSTRACT**

Described herein are fusion proteins that include an intein polypeptide along with methods of using of such proteins. Fusion proteins described herein may include a Cas12a polypeptide and an intein polypeptide or a polypeptide of interest (e.g., a reverse transcriptase polypeptide) and an intein polypeptide. Also described herein are compositions and systems for modifying or editing a target nucleic acid.

Specification includes a Sequence Listing.

FIG. 1

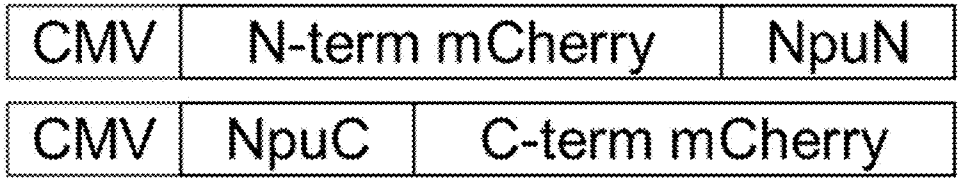

CMV
N-term mCherry
NpuN
CMV
NpuC
C-term mCherry

FIG. 2

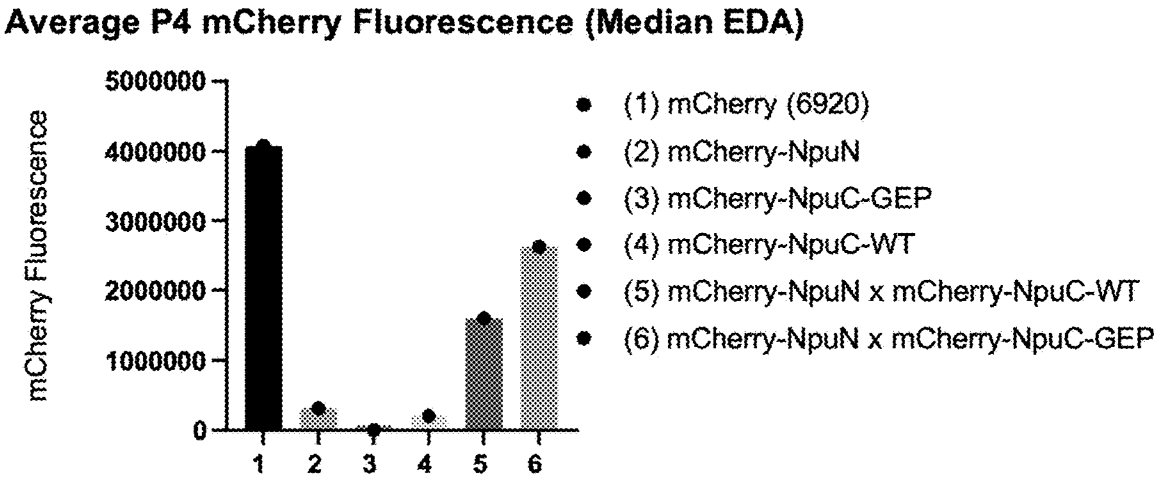

Average P4 mCherry Fluorescence (Median EDA)
mCherry Fluorescence
5000000
4000000
3000000
2000000
1000000
0
1
2
3
4
5
6
(1) mCherry (6920)
(2) mCherry-NpuN
(3) mCherry-NpuC-GEP
(4) mCherry-NpuC-WT
(5) mCherry-NpuN x mCherry-NpuC-WT
(6) mCherry-NpuN x mCherry-NpuC-GEP

FIG. 3

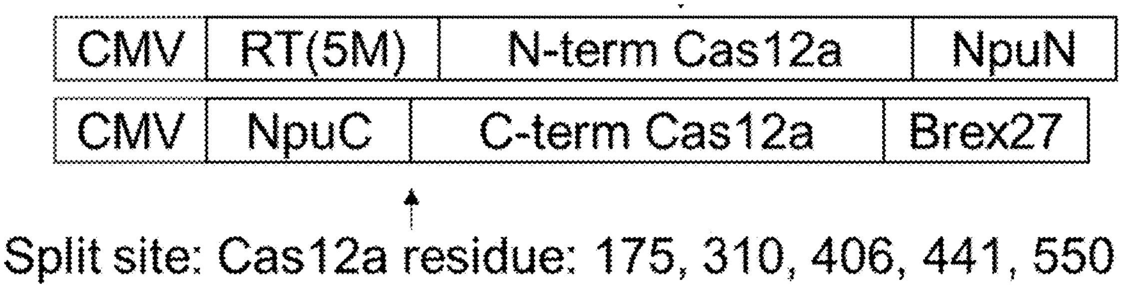

CMV
RT(5M)
N-term Cas12a
NpuN
CMV
NpuC
C-term Cas12a
Brex27
Split site: Cas12a residue: 175, 310, 406, 441, 550

FIG. 4

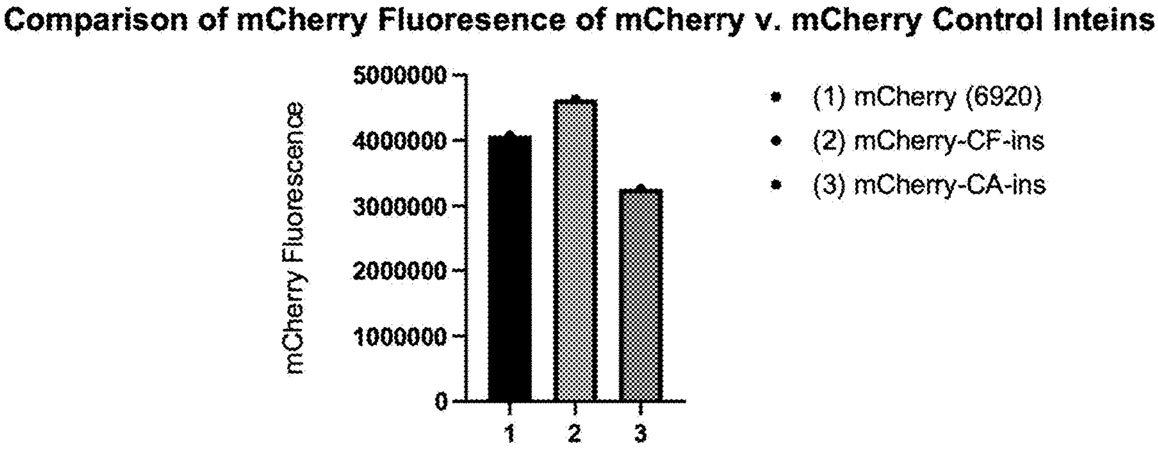

Comparison of mCherry Fluoresence of mCherry v. mCherry Control Inteins
mCherry Fluorescence
5000000
4000000
3000000
2000000
1000000
0
1
2
3
(1) mCherry (6920)
(2) mCherry-CF-ins
(3) mCherry-CA-ins

FIG. 5

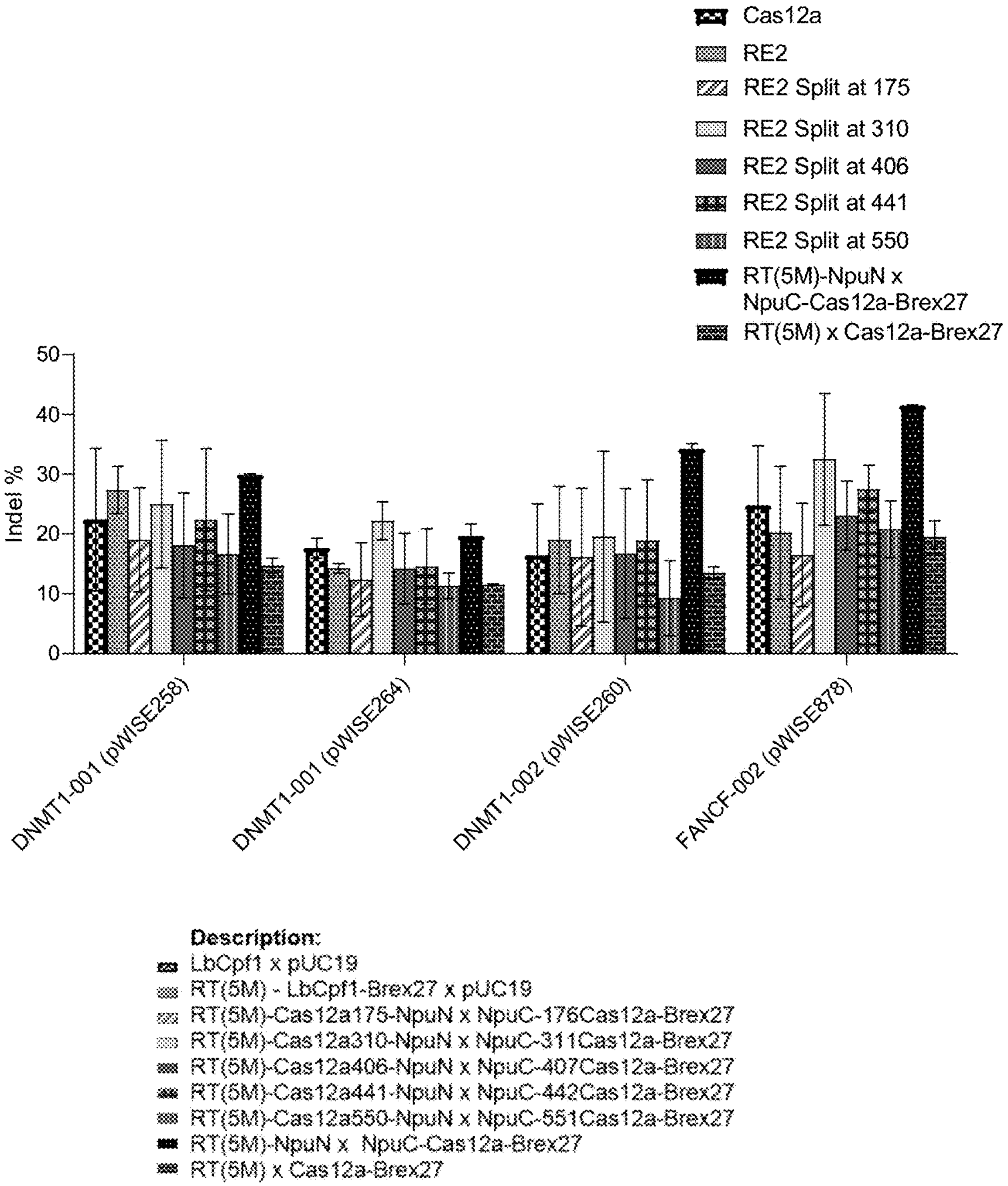
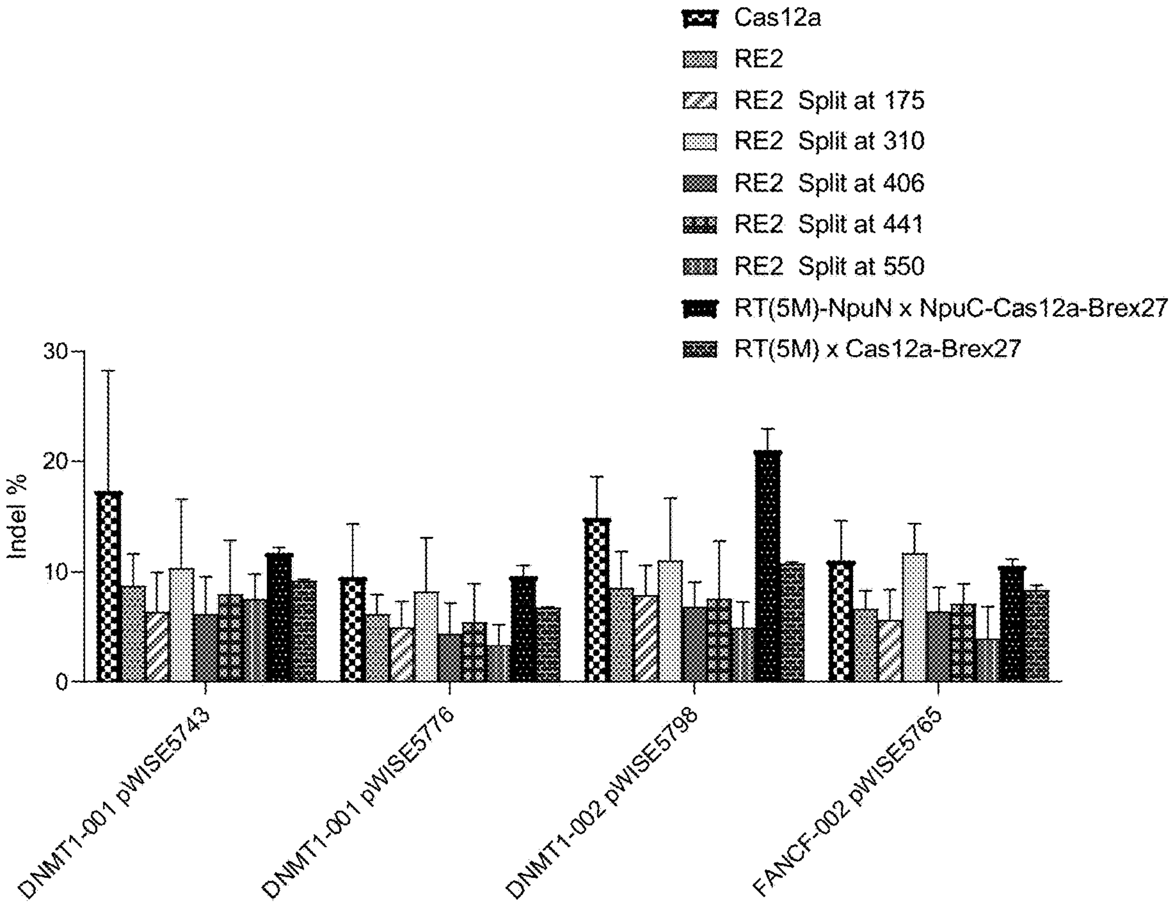

Cas12a
RE2
RE2 Split at 175
RE2 Split at 310
RE2 Split at 406
RE2 Split at 441
RE2 Split at 550
RT(5M)-NpuN x NpuC-Cas12a-Brex27
RT(5M) x Cas12a-Brex27
Indel %
0
10
20
30
40
50
DNMT1-001 (pWISE258)
DNMT1-001 (pWISE264)
DNMT1-002 (pWISE260)
FANCF-002 (pWISE878)
Description:
LbCpf1 x pUC19
RT(5M) - LbCpf1-Brex27 x pUC19
RT(5M)-Cas12a175-NpuN x NpuC-176Cas12a-Brex27
RT(5M)-Cas12a310-NpuN x NpuC-311Cas12a-Brex27
RT(5M)-Cas12a406-NpuN x NpuC-407Cas12a-Brex27
RT(5M)-Cas12a441-NpuN x NpuC-442Cas12a-Brex27
RT(5M)-Cas12a550-NpuN x NpuC-551Cas12a-Brex27
RT(5M)-NpuN x NpuC-Cas12a-Brex27
RT(5M) x Cas12a-Brex27

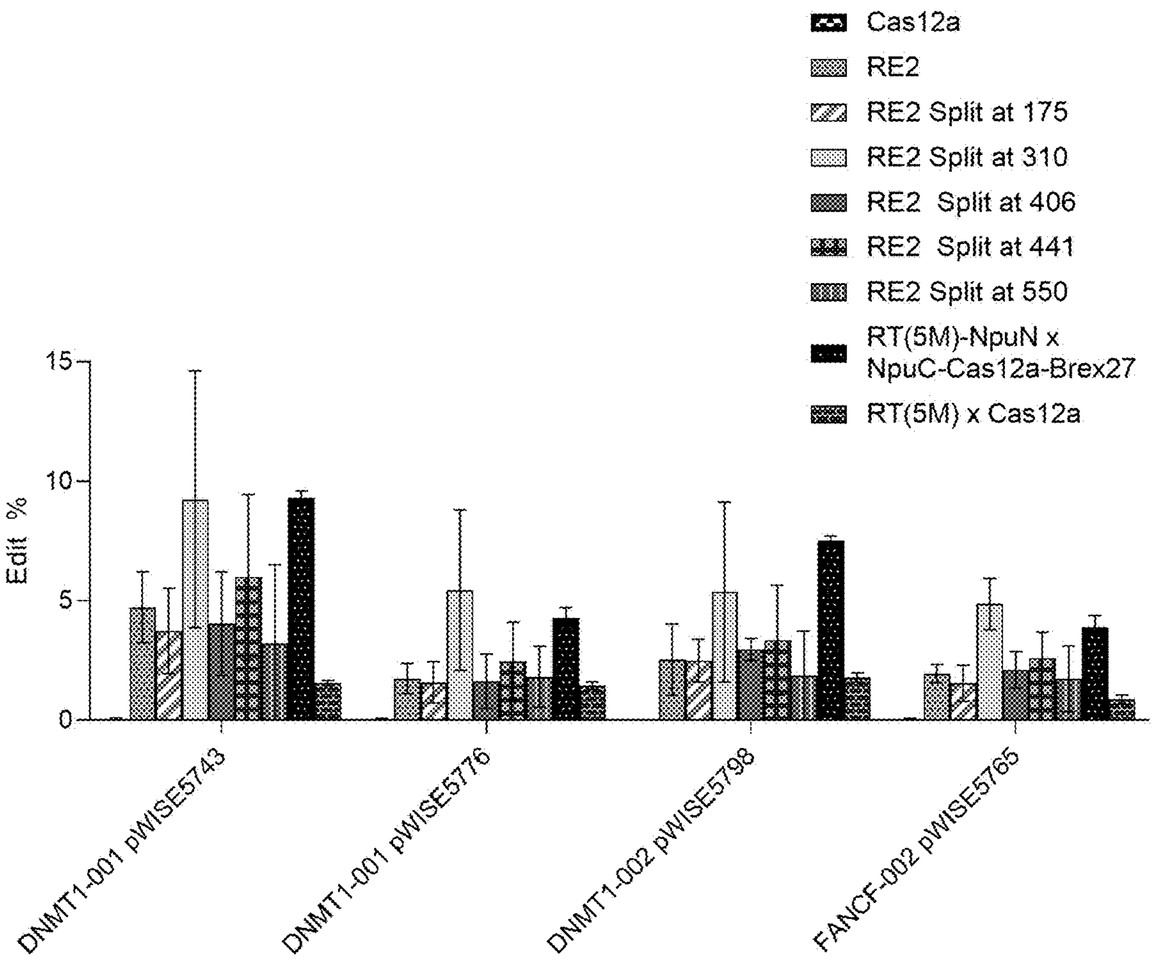
FIG. 6



Description:

- ▨ LbCpf1 x pUC19
- ▨ RT(5M) - LbCpf1-Brex27 x pUC19
- ▨ RT(5M)-Cas12a175-NpuN x NpuC-176Cas12a-Brex27
- ▨ RT(5M)-Cas12a310-NpuN x NpuC-311Cas12a-Brex27
- ▨ RT(5M)-Cas12a406-NpuN x NpuC-407Cas12a-Brex27
- ▨ RT(5M)-Cas12a441-NpuN x NpuC-442Cas12a-Brex27
- ▨ RT(5M)-Cas12a550-NpuN x NpuC-551Cas12a-Brex27
- ■ RT(5M)-NpuN x NpuC-Cas12a-Brex27
- ▨ RT(5M) x Cas12a-Brex27

FIG. 7



Description:

- ▨ LbCpf1 x pUC19
- ▨ RT(5M) - LbCpf1-Brex27 x pUC19
- ▨ RT(5M)-Cas12a175-NpuN x NpuC-176Cas12a-Brex27
- ▨ RT(5M)-Cas12a310-NpuN x NpuC-311Cas12a-Brex27
- ▨ RT(5M)-Cas12a406-NpuN x NpuC-407Cas12a-Brex27
- ▨ RT(5M)-Cas12a441-NpuN x NpuC-442Cas12a-Brex27
- ▨ RT(5M)-Cas12a550-NpuN x NpuC-551Cas12a-Brex27
- ■ RT(5M)-NpuN x NpuC-Cas12a-Brex27
- ▨ RT(5M) x Cas12a-Brex27

FUSION PROTEINS COMPRISING AN INTEIN POLYPEPTIDE AND METHODS OF USE THEREOF

STATEMENT REGARDING ELECTRONIC FILING OF A SEQUENCE LISTING

[0001] A Sequence Listing in XML format, entitled 1499-116_ST26.xml, 502,046 bytes in size, generated on Dec. 14, 2023, and filed herewith, is hereby incorporated by reference in its entirety for its disclosures.

FIELD

[0002] This invention relates to fusion proteins (e.g., engineered proteins) that include a Cas12a polypeptide and an intein polypeptide and to methods of use of such proteins. The invention also relates to fusion proteins (e.g., engineered proteins) that include a polypeptide of interest (e.g., a reverse transcriptase polypeptide) and an intein polypeptide and to methods of use of such proteins. The invention further relates to compositions and systems for modifying or editing a target nucleic acid.

BACKGROUND OF THE INVENTION

[0003] Large genome editing agents (e.g., those over 2000 amino acids in length) are usually delivered into tissues using means other than Adeno-associated virus (AAV) vectors. For example, large genome editing agents may be delivered using lipid nanoparticle mediated RNP delivery, which has no size limit, or using mRNA or DNA. However, these approaches prohibit use in applications where the use of AAV is necessary, for example, applications for delivery to the brain of a subject.

[0004] Accordingly, new methods for preparing and/or delivering genome editing agents are needed.

SUMMARY OF THE INVENTION

[0005] A first aspect of the present invention is directed to a fusion protein comprising an intein polypeptide. In some embodiments, the fusion protein comprises a Cas12a polypeptide fused to an intein polypeptide. In some embodiments, the fusion protein comprises a polypeptide of interest fused to an intein polypeptide. In some embodiments, the fusion protein comprises a reverse transcriptase polypeptide fused to an intein polypeptide. A nucleic acid molecule encoding a fusion protein as described herein is also provided.

[0006] Another aspect of the present invention is directed to a complex comprising: a Cas12a protein that is prepared from a first fusion protein and a second fusion protein, wherein the first fusion protein comprises a first Cas12a polypeptide fused to a first intein polypeptide and the second fusion protein comprises a second Cas12a polypeptide fused to a second intein polypeptide; a guide nucleic acid (e.g., a guide RNA); and optionally a deaminase.

[0007] An additional aspect of the present invention is directed to a complex comprising: an engineered protein (e.g., a base editor or a templated editor such as a REDRAW editor) that is prepared from a first fusion protein and a second fusion protein, wherein the first fusion protein comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide and the second fusion protein comprises a Cas12a polypeptide fused to a second intein polypeptide; and a guide nucleic acid (e.g., a guide RNA).

[0008] A further aspect of the present invention is directed to a method of modifying a target nucleic acid, the method comprising: contacting the target nucleic acid with: a Cas12a protein that is prepared from a first fusion protein and a second fusion protein, wherein the first fusion protein comprises a first Cas12a polypeptide fused to a first intein polypeptide and the second fusion protein comprises a second Cas12a polypeptide fused to a second intein polypeptide; and a guide nucleic acid (e.g., a guide RNA), optionally wherein the Cas12a protein and the guide nucleic acid form a complex or are comprised in a complex, thereby modifying the target nucleic acid.

[0009] Another aspect of the present invention is directed to a method of modifying a target nucleic acid, the method comprising: contacting the target nucleic acid with: an engineered protein (e.g., a base editor or a templated editor such as a REDRAW editor) that is prepared from a first fusion protein and a second fusion protein, wherein the first fusion protein comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide and the second fusion protein comprises a Cas12a polypeptide fused to a second intein polypeptide; and a guide nucleic acid (e.g., a guide RNA), optionally wherein the engineered protein and the guide nucleic acid form a complex or are comprised in a complex, thereby modifying the target nucleic acid.

[0010] An additional aspect of the present invention is directed to a composition comprising: a first fusion protein that comprises a first Cas12a polypeptide fused to a first intein polypeptide; and a second fusion protein that comprises a second Cas12a polypeptide fused to a second intein polypeptide.

[0011] A further aspect of the present invention is directed to a composition comprising: a first fusion protein that comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide; and a second fusion protein that comprises a Cas12a polypeptide fused to a second intein polypeptide.

[0012] Another aspect of the present invention is directed to a composition comprising: a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide.

[0013] An additional aspect of the present invention is directed to a composition comprising: a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide.

[0014] A further aspect of the present invention is directed to a kit comprising: a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide.

[0015] Another aspect of the present invention is directed to a kit comprising: a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide.

[0016] An additional aspect of the present invention is directed to a method of modifying a target nucleic acid, the method comprising: introducing a first nucleic acid molecule and a second nucleic acid molecule into a cell, wherein the first nucleic acid molecule encodes a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide and the second nucleic acid molecule encodes a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide; contacting the target nucleic acid in the cell with a protein comprising at least a portion of the first Cas12a polypeptide and at least a portion of the second Cas12a polypeptide and a guide nucleic acid (e.g., a guide RNA), optionally wherein the protein and the guide nucleic acid form a complex or are comprised in a complex, thereby modifying the target nucleic acid.

[0017] Another aspect of the present invention is directed to a method of modifying a target nucleic acid, the method comprising: introducing a first nucleic acid molecule and a second nucleic acid molecule into a cell, wherein the first nucleic acid molecule encodes a first fusion protein, the first fusion protein comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide and the second nucleic acid molecule encodes a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide; contacting the target nucleic acid in the cell with a protein comprising at least a portion of the polypeptide of interest and at least a portion of the Cas12a polypeptide and a guide nucleic acid (e.g., a guide RNA), optionally wherein the protein and the guide nucleic acid form a complex or are comprised in a complex, thereby modifying the target nucleic acid.

[0018] The invention further provides expression cassettes and/or vectors comprising a nucleic acid construct of the present invention, and cells comprising a polypeptide, fusion protein and/or nucleic acid construct of the present invention. Additionally, the invention provides kits comprising a nucleic acid construct of the present invention and expression cassettes, vectors and/or cells comprising the same.

[0019] It is noted that aspects of the invention described with respect to one embodiment, may be incorporated in a different embodiment although not specifically described relative thereto. That is, all embodiments and/or features of any embodiment can be combined in any way and/or combination. Applicant reserves the right to change any originally filed claim and/or file any new claim accordingly, including the right to be able to amend any originally filed claim to depend from and/or incorporate any feature of any other claim or claims although not originally claimed in that manner. These and other objects and/or aspects of the present invention are explained in detail in the specification set forth below. Further features, advantages and details of the present invention will be appreciated by those of ordinary skill in the art from a reading of the figures and the detailed description of the preferred embodiments that follow, such description being merely illustrative of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] FIG. **1** is a schematic illustration of two exemplary split proteins for mCherry.

[0021] FIG. **2** is a graph showing the average mCherry fluorescence for different proteins, which can measure the intein splicing activity for a mCherry reconstitution assay. (1): is the expression of wildtype full length mCherry. (2): is only expressing N-terminal mCherry fused to N-terminal Npu split intein. (3): is only expressing C-terminal mCherry fused to C-terminal Npu split intein with the GEP mutation that enhances intein activity and robustness. (4) is only expressing C-terminal mCherry fused to wildtype (WT) C-terminal Npu split intein. (5) is expressing both N- and C-terminal mCherry parts, using the wildtype variant of the Npu intein. (6) is expressing both N- and C-terminal mCherry parts, using the GEP variant of the Npu intein.

[0022] FIG. **3** is a schematic illustration of two split proteins for Redraw Editor 2 (RE2) comprising Cas12a showing exemplary split sites according to some embodiments of the present invention.

[0023] FIG. **4** is a graph showing the effect of a two amino acid insertion (e.g., a CF or CA residues) on mCherry fluorescence.

[0024] FIG. **5** is a graph showing the percentage of indels generated using non-split and split Redraw editors using crRNAs according to some embodiments of the present invention.

[0025] FIG. **6** is a graph showing the percentage of indels generated using non-split and split Redraw editors using stagRNAs according to some embodiments of the present invention.

[0026] FIG. **7** is a graph showing the percentage of precise edits generated using non-split and split Redraw editors using stagRNAs according to some embodiments of the present invention.

DETAILED DESCRIPTION

[0027] The present invention now will be described hereinafter with reference to the accompanying drawings and examples, in which embodiments of the invention are shown. This description is not intended to be a detailed catalog of all the different ways in which the invention may be implemented, or all the features that may be added to the instant invention. For example, features illustrated with respect to one embodiment may be incorporated into other embodiments, and features illustrated with respect to a particular embodiment may be deleted from that embodiment. Thus, the invention contemplates that in some embodiments of the invention, any feature or combination of features set forth herein can be excluded or omitted. In addition, numerous variations and additions to the various embodiments suggested herein will be apparent to those skilled in the art in light of the instant disclosure, which do not depart from the instant invention. Hence, the following descriptions are intended to illustrate some particular embodiments of the invention, and not to exhaustively specify all permutations, combinations and variations thereof.

[0028] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention.

[0029] All publications, patent applications, patents and other references cited herein are incorporated by reference in their entireties for the teachings relevant to the sentence and/or paragraph in which the reference is presented.

[0030] Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination. Moreover, the present invention also contemplates that in some embodiments of the invention, any feature or combination of features set forth herein can be excluded or omitted. To illustrate, if the specification states that a composition comprises components A, B and C, it is specifically intended that any of A, B or C, or a combination thereof, can be omitted and disclaimed singularly or in any combination.

[0031] As used in the description of the invention and the appended claims, the singular forms "a," "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

[0032] Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative ("or").

[0033] The term "about," as used herein when referring to a measurable value such as an amount or concentration and the like, is meant to encompass variations of ±10%, ±5%, ±1%, ±0.5%, or even ±0.1% of the specified value as well as the specified value. For example, "about X" where X is the measurable value, is meant to include X as well as variations of ±10%, ±5%, ±1%, ±0.5%, or even ±0.1% of X. A range provided herein for a measurable value may include any other range and/or individual value therein.

[0034] As used herein, phrases such as "between X and Y" and "between about X and Y" should be interpreted to include X and Y. As used herein, phrases such as "between about X and Y" mean "between about X and about Y" and phrases such as "from about X to Y" mean "from about X to about Y."

[0035] Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. For example, if the range 10 to 15 is disclosed, then 11, 12, 13, and 14 are also disclosed.

[0036] The term "comprise," "comprises" and "comprising" as used herein, specify the presence of the 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.

[0037] As used herein, the transitional phrase "consisting essentially of" means that the scope of a claim is to be interpreted to encompass the specified materials or steps recited in the claim and those that do not materially affect the basic and novel characteristic(s) of the claimed invention. Thus, the term "consisting essentially of" when used in a claim of this invention is not intended to be interpreted to be equivalent to "comprising."

[0038] As used herein, the terms "increase," "increasing," "enhance," "enhancing," "improve" and "improving" (and grammatical variations thereof) describe an elevation of at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 150%, 200%, 300%, 400%, 500% or more such as compared to another measurable property or quantity (e.g., a control value).

[0039] As used herein, the terms "reduce," "reduced," "reducing," "reduction," "diminish," and "decrease" (and grammatical variations thereof), describe, for example, a decrease of at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% such as compared to another measurable property or quantity (e.g., a control value). In some embodiments, the reduction can result in no or essentially no (i.e., an insignificant amount, e.g., less than about 10% or even 5%) detectable activity or amount.

[0040] A "heterologous nucleotide sequence" or a "recombinant nucleotide sequence" is a nucleotide sequence not naturally associated with a host cell into which it is introduced, including non-naturally occurring multiple copies of a naturally occurring nucleotide sequence.

[0041] A "native" or "wild-type" nucleic acid, nucleotide sequence, polypeptide or amino acid sequence refers to a naturally occurring or endogenous nucleic acid, nucleotide sequence, polypeptide or amino acid sequence. Thus, for example, a "native nucleic acid" is a nucleic acid that is naturally occurring in or endogenous to a reference organism. A "homologous" nucleic acid sequence is a nucleotide sequence naturally associated with a host cell into which it is introduced.

[0042] As used herein, the terms "nucleic acid," "nucleic acid molecule," "nucleotide sequence" and "polynucleotide" refer to RNA or DNA that is linear or branched, single or double stranded, or a hybrid thereof. The term also encompasses RNA/DNA hybrids. When dsRNA is produced synthetically, less common bases, such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others can also be used for antisense, dsRNA, and ribozyme pairing. For example, polynucleotides that contain C-5 propyne analogues of uridine and cytidine have been shown to bind RNA with high affinity and to be potent antisense inhibitors of gene expression. Other modifications, such as modification to the phosphodiester backbone, or the 2'-hydroxy in the ribose sugar group of the RNA can also be made.

[0043] As used herein, the term "nucleotide sequence" refers to a heteropolymer of nucleotides or the sequence of these nucleotides from the 5' to 3' end of a nucleic acid molecule and includes DNA or RNA molecules, including cDNA, a DNA fragment or portion, genomic DNA, synthetic (e.g., chemically synthesized) DNA, plasmid DNA, mRNA, and anti-sense RNA, any of which can be single stranded or double stranded. The terms "nucleotide sequence" "nucleic acid," "nucleic acid molecule," "nucleic acid construct," "recombinant nucleic acid," "oligonucleotide" and "polynucleotide" are also used interchangeably herein to refer to a heteropolymer of nucleotides. Nucleic acid molecules and/or nucleotide sequences provided herein are presented herein in the 5' to 3' direction, from left to right and are represented using the standard code for representing the nucleotide characters as set forth in the U.S. sequence rules, 37 CFR §§ 1.821-1.825 and the World Intellectual Property Organization (WIPO) Standard ST.25. A "5' region" as used herein can mean the region of a polynucleotide that is nearest the 5' end of the polynucleotide. Thus, for example, an element in the 5' region of a polynucleotide can be located anywhere from the first nucleotide located at the 5' end of the polynucleotide to the nucleotide located halfway through the polynucleotide. A "3' region" as used herein can mean the region of a polynucleotide that is nearest the 3' end of the polynucleotide. Thus, for example, an element in the 3' region of a polynucleotide can be located anywhere from the first nucleotide located at the 3' end of the polynucleotide to the nucleotide located halfway through the polynucleotide.

[0044] As used herein, the term "gene" refers to a nucleic acid molecule capable of being used to produce mRNA, antisense RNA, miRNA, anti-microRNA antisense oligodeoxyribonucleotide (AMO) and the like. Genes may or may not be capable of being used to produce a functional protein or gene product. Genes can include both coding and non-coding regions (e.g., introns, regulatory elements, promoters, enhancers, termination sequences and/or 5' and 3' untranslated regions).

[0045] A polynucleotide, gene, or polypeptide may be "isolated" by which is meant a nucleic acid or polypeptide that is substantially or essentially free from components normally found in association with the nucleic acid or polypeptide, respectively, in its natural state. In some embodiments, such components include other cellular material, culture medium from recombinant production, and/or various chemicals used in chemically synthesizing the nucleic acid or polypeptide.

[0046] The term "mutation" refers to point mutations (e.g., missense, or nonsense, or insertions or deletions of single base pairs that result in frame shifts), insertions, deletions, and/or truncations. When the mutation is a substitution of a residue within an amino acid sequence with another residue, or a deletion or insertion of one or more residues within a sequence, the mutations are typically described by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue.

[0047] The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" (5' to 3') binds to the complementary sequence "T-C-A" (3' to 5'). Complementarity between two single-stranded molecules may be "partial," in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

[0048] "Complement" as used herein can mean 100% complementarity with the comparator nucleotide sequence or it can mean less than 100% complementarity (e.g., "substantially complementary," such as about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and the like, complementarity).

[0049] A "portion" or "fragment" of a nucleotide sequence or polypeptide (including a domain) will be understood to mean a nucleotide sequence or polypeptide of reduced length (e.g., reduced by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more residue(s) (e.g., nucleotide(s) or peptide(s)) relative to a reference nucleotide sequence or polypeptide, respectively, and comprising, consisting essentially of and/or consisting of a nucleotide sequence or polypeptide of contiguous residues, respectively, identical or almost identical (e.g., 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical) to the reference nucleotide sequence or polypeptide. In some embodiments, a portion of a reference nucleotide sequence or polypeptide is about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or more of the full-length reference nucleotide sequence or polypeptide. Such a nucleic acid fragment or portion according to the invention may be, where appropriate, included in a larger polynucleotide of which it is a constituent. As an example, a repeat sequence of guide nucleic acid of this invention may comprise a portion of a wild-type CRISPR-Cas repeat sequence (e.g., a wild-type Type V CRISPR Cas repeat, e.g., a repeat from the CRISPR Cas system that includes, but is not limited to, Cas12a (Cpf1), Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c1, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, and/or Cas14c, and the like). Similarly a portion of a polypeptide may be included in a larger polypeptide of which it is a constituent.

[0050] Different nucleic acids or proteins having homology are referred to herein as "homologues." The term homologue includes homologous sequences from the same and other species and orthologous sequences from the same and other species. "Homology" refers to the level of similarity between two or more nucleic acid and/or amino acid sequences in terms of percent of positional identity (i.e., sequence similarity or identity). Homology also refers to the concept of similar functional properties among different nucleic acids or proteins. Thus, the compositions and methods of the invention further comprise homologues to the nucleotide sequences and polypeptides of this invention. "Orthologous" and "orthologs" as used herein, refers to homologous nucleotide sequences and/or amino acid sequences in different species that arose from a common ancestral gene during speciation. A homologue or ortholog of a nucleotide sequence of this invention has a substantial sequence identity (e.g., at least about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or 100%) to said nucleotide sequence of the invention.

[0051] As used herein "sequence identity" refers to the extent to which two optimally aligned polynucleotide or polypeptide sequences are invariant throughout a window of alignment of components, e.g., nucleotides or amino acids. "Identity" can be readily calculated by known methods including, but not limited to, those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data,* *Part I* (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); and *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991).

[0052] As used herein, the term "percent sequence identity" or "percent identity" refers to the percentage of identical nucleotides in a linear polynucleotide sequence of a reference ("query") polynucleotide molecule (or its complementary strand) as compared to a test ("subject") polynucleotide molecule (or its complementary strand) when the two sequences are optimally aligned. In some embodiments, "percent identity" can refer to the percentage of identical amino acids in an amino acid sequence as compared to a reference polypeptide.

[0053] As used herein, the phrase "substantially identical," or "substantial identity" in the context of two nucleic acid molecules, nucleotide sequences or protein sequences, refers to two or more sequences or subsequences that have at least about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or 100% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. In some embodiments of the invention, the substantial identity exists over a region of consecutive nucleotides of a nucleotide sequence of the invention that is about 10 nucleotides to about 20 nucleotides, about 10 nucleotides to about 25 nucleotides, about 10 nucleotides to about 30 nucleotides, about 15 nucleotides to about 25 nucleotides, about 30 nucleotides to about 40 nucleotides, about 50 nucleotides to about 60 nucleotides, about 70 nucleotides to about 80 nucleotides, about 90 nucleotides to about 100 nucleotides, or more nucleotides in length, and any range therein, up to the full length of the sequence. In some embodiments, the nucleotide sequences can be substantially identical over at least about 20 nucleotides (e.g., about 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 nucleotides). In some embodiments, a substantially identical nucleotide or protein sequence performs substantially the same function as the nucleotide (or encoded protein sequence) to which it is substantially identical.

[0054] For sequence comparison, typically one sequence acts as a reference sequence to which one or more test sequence(s) are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0055] Optimal alignment of sequences for aligning a comparison window are well known to those skilled in the art and may be conducted by tools such as the local homology algorithm of Smith and Waterman, the homology alignment algorithm of Needleman and Wunsch, the search for similarity method of Pearson and Lipman, and optionally by computerized implementations of these algorithms such as GAP, BESTFIT, FASTA, and TFASTA available as part of the GCG® Wisconsin Package® (Accelrys Inc., San Diego, CA) as well as web-based alignment programs such as Clustal Omega, EMBOSS Needle, EMBOSS Stretcher, EMBOSS Water, LALIGN, GGSEARCH2SEQ, EMBOS Cons, Kalign, MAFFT, MUSCLE, and T-Coffee. In some embodiments, an "optimal alignment" of two sequences (e.g., two polypeptide sequences) is the highest scoring alignment, optionally from an alignment conducted by a tool such as the local homology algorithm of Smith and Waterman, the homology alignment algorithm of Needleman and Wunsch, the search for similarity method of Pearson and Lipman, GAP, BESTFIT, FASTA, and TFASTA available as part of the GCG® Wisconsin Package® (Accelrys Inc., San Diego, CA), Clustal Omega, EMBOSS Needle, EMBOSS Stretcher, EMBOSS Water, LALIGN, GGSEARCH2SEQ, EMBOS Cons, Kalign, MAFFT, MUSCLE, and/or T-Coffee. In some embodiments, an "optimal alignment" of two sequences (e.g., two polypeptide sequences) is the alignment that provides the highest percent sequence identity, optionally allowing for one or more gap(s) to be introduced into one or both sequences. An "identity fraction" for aligned segments of a test sequence and a reference sequence is the number of identical components which are shared by the two aligned sequences divided by the total number of components in the reference sequence segment, e.g., the entire reference sequence or a smaller defined part of the reference sequence. Percent sequence identity is represented as the identity fraction multiplied by 100. The comparison of one or more sequence(s) may be to a full-length sequence or a portion thereof, or to a longer sequence. For purposes of this invention "percent identity" and/or optimal alignment may be determined using Basic Local Alignment Search Tool (BLAST) provided by the National Center for Biotechnology Information such as BLASTX, for translated nucleotide sequences, BLASTN for polynucleotide sequences, and BLASTP for polypeptide sequences.

[0056] Two nucleotide sequences may also be considered substantially complementary when the two sequences hybridize to each other under stringent conditions. In some representative embodiments, two nucleotide sequences considered to be substantially complementary hybridize to each other under highly stringent conditions.

[0057] "Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993). Generally, highly stringent hybridization and wash conditions are selected to be about 5° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

[0058] The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleotide sequences which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42° C., with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.1 5M NaCl at 72° C. for about 15 minutes. An example of stringent wash conditions is a 0.2×SSC wash at 65° C. for 15 minutes (see, Sambrook, infra, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example of a medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1×SSC at 45° C. for 15 minutes. An example of a low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6×SSC at 40° C. for 15 minutes. For short probes (e.g., about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30° C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2× (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleotide sequences that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This can occur, for example, when a copy of a nucleotide sequence is created using the maximum codon degeneracy permitted by the genetic code.

[0059] A polynucleotide and/or recombinant nucleic acid construct of this invention can be codon optimized for expression. In some embodiments, a polynucleotide, nucleic acid construct, expression cassette, and/or vector of the present invention (e.g., that comprises/encodes a fusion protein, a nucleic acid binding polypeptide (e.g., a DNA binding polypeptide such as a sequence-specific DNA binding domain from a polynucleotide-guided endonuclease, a zinc finger nuclease, a transcription activator-like effector nuclease (TALEN), an Argonaute protein, and/or a CRISPR-Cas effector protein), a guide nucleic acid, a cytosine deaminase and/or adenine deaminase) may be codon optimized for expression in an organism (e.g., an animal such as a human, a plant, a fungus, an archaeon, or a bacterium). In some embodiments, the codon optimized nucleic acid constructs, polynucleotides, expression cassettes, and/or vectors of the invention have about 70% to about 99.9% (e.g., 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%. 99.9% or 100%) identity or more to the reference nucleic acid constructs, polynucleotides, expression cassettes, and/or vectors but which have not been codon optimized.

[0060] In any of the embodiments described herein, a polynucleotide or nucleic acid construct of the invention may be operatively associated with a variety of promoters and/or other regulatory elements for expression in an organism or cell thereof (e.g., a mammal and/or a mammalian cell, a plant and/or a cell of a plant, etc.). Thus, in some embodiments, a polynucleotide or nucleic acid construct of this invention may further comprise one or more promoters, introns, enhancers, and/or terminators operably linked to one or more nucleotide sequences. In some embodiments, a promoter may be operably associated with an intron (e.g., Ubi1 promoter and intron). In some embodiments, a promoter associated with an intron maybe referred to as a "promoter region" (e.g., Ubi1 promoter and intron).

[0061] By "operably linked" or "operably associated" as used herein in reference to polynucleotides, it is meant that the indicated elements are functionally related to each other, and are also generally physically related. Thus, the term "operably linked" or "operably associated" as used herein, refers to nucleotide sequences on a single nucleic acid molecule that are functionally associated. Thus, a first nucleotide sequence that is operably linked to a second nucleotide sequence means a situation when the first nucleotide sequence is placed in a functional relationship with the second nucleotide sequence. For instance, a promoter is operably associated with a nucleotide sequence if the promoter effects the transcription or expression of said nucleotide sequence. Those skilled in the art will appreciate that the control sequences (e.g., promoter) need not be contiguous with the nucleotide sequence to which it is operably associated, as long as the control sequences function to direct the expression thereof. Thus, for example, intervening untranslated, yet transcribed, nucleic acid sequences can be present between a promoter and the nucleotide sequence, and the promoter can still be considered "operably linked" to the nucleotide sequence.

[0062] As used herein, the term "linked," or "fused" in reference to polypeptides, refers to the covalent attachment of one polypeptide to another. A polypeptide may be linked or fused to another polypeptide (e.g., at the N-terminus or the C-terminus) directly (e.g., via a peptide bond) or through a linker (e.g., a peptide linker). Two polypeptides being directly fused (e.g., a direct linkage) refers to the covalent attachment of one amino acid residue of a first polypeptide of the two polypeptides to an amino acid residue of a second polypeptide of the two polypeptides without an intervening element between the two amino acid residues. For example, first and second polypeptides may be directly linked via a peptide bond between the first and second polypeptides without an intervening element (e.g., a linker) between the first and second polypeptides. Two polypeptides being indirectly fused (e.g., an indirect linkage) refers to an intervening element (e.g., a linker such as a peptide linker) that is present between the two polypeptides and is covalently attached to each, optionally the intervening element may attach one end of a first polypeptide of the two polypeptides to an end of the second polypeptide of the two polypeptides.

[0063] A "fusion protein" as used herein refers to two or more polypeptides that are covalently attached (e.g., directly or indirectly) so that they are transcribed and translated as a single unit and thereby produce a single polypeptide comprising the two or more polypeptides. In some embodiments, the two or more polypeptides may naturally be encoded by separate genes, but, in the form of a fusion protein, are encoded by a single gene.

[0064] The term "linker" is art-recognized and refers to a chemical group or a molecule linking two molecules or moieties, e.g., linking two polypeptides or domains of a fusion protein, such as, for example, a Cas12a polypeptide and an intein polypeptide. A linker may be comprised of a single linking molecule (e.g., a single amino acid) or may comprise more than one linking molecule. In some embodiments, the linker can be an organic molecule, group, polymer, or chemical moiety such as a bivalent organic moiety. In some embodiments, the linker may be an amino acid or it may be a peptide. In some embodiments, the linker is a peptide (e.g., a peptide linker).

[0065] In some embodiments, a peptide linker useful with this invention may be about 2 to about 100 or more amino acids in length, for example, about 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, 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length (e.g., about 2 to about 40, about 2 to about 50, about 2 to about 60, about 4 to about 40, about 4 to about 50, about 4 to about 60, about 5 to about 40, about 5 to about 50, about 5 to about 60, about 9 to about 40, about 9 to about 50, about 9 to about 60, about 10 to about 40, about 10 to about 50, about 10 to about 60, or about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 amino acids to about 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more amino acids in length (e.g., about 105, 110, 115, 120, 130, 140 150 or more amino acids in length)). In some embodiments, a peptide linker may comprise glycine (G) and serine (S) such as a GS linker. In some embodiments, a peptide linker may comprise a cysteine (C) and alanine (A) such as a CA linker. In some embodiments, a peptide linker may comprise a cysteine (C) and phenylalanine (F) such as a CF linker. In some embodiments, the peptide linker is a GS linker, a CA linker, or a CF linker having 2, 3, or 4 amino acid residues, optionally 2 or 4 amino acid residues. In some embodiments, the peptide linker has one of the amino acid sequences of SEQ ID NOs:1-35. In some embodiments, the peptide linker may comprise an amino acid sequence of CA, CF, $(GGS)_n$, GS, SG, GSSG (SEQ ID NO:31), GSSGSS (SEQ ID NO:32), GSSGSSGS (SEQ ID NO:33), $(GSS)_n$ (SEQ ID NO:34), $(GSS)_n$GS (SEQ ID NO:35), S$(GGS)_n$ (SEQ ID NO:25), SGGS (SEQ ID NO:26), (GSS)nG (SEQ ID NO:191), or (GGGGS)n (SEQ ID NO:27), wherein n is an integer of 1-20 (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20). In some embodiments, the peptide linker may comprise the amino acid sequence: SGGSGGSGGS (SEQ ID NO:28). In some embodiments, the peptide linker may comprise the amino acid sequence: SGSETPGTSESATPES (SEQ ID NO:29), also referred to as the XTEN linker. In some embodiments, the peptide linker may comprise the amino acid sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO:30), also referred to as the GS-XTEN-GS linker. In some embodiments, a peptide linker has an amino acid sequence of SEQ ID NOs:189 or SEQ ID NOs:190.

[0066] As used herein, the term "linked," or "fused" in reference to polynucleotides, refers to the covalent attachment of one polynucleotide to another polynucleotide. In some embodiments, two or more polynucleotide molecules may be linked by a linker that can be an organic molecule, group, polymer, or chemical moiety such as a bivalent organic moiety. A polynucleotide may be linked or fused to another polynucleotide (at the 5' end or the 3' end) via a direct covalent linkage or through one or more linking nucleotides. In some embodiments, a polynucleotide motif of a certain structure may be inserted within another polynucleotide sequence (e.g., extension of the hairpin structure in guide RNA). In some embodiments, the linking nucleotides may be naturally occurring nucleotides. In some embodiments, the linking nucleotides may be non-naturally occurring nucleotides. Two polynucleotides being directly fused (e.g., a direct linkage) refers to the covalent attachment of one nucleotide of a first polynucleotide of the two polynucleotides to a nucleotide of a second polynucleotide of the two polynucleotides without an intervening element between the two polynucleotides. For example, first and second polynucleotides may be directly linked via a phosphodiester bond between the first and second polynucleotides without an intervening element (e.g., a linker) between the first and second polynucleotides. Two polynucleotides being indirectly fused (e.g., an indirect linkage) refers to an intervening element (e.g., a linker such as a polynucleotide linker) that is present between the two polynucleotides and is covalently attached to each, optionally the intervening element attaches one end of a first polynucleotide of the two polynucleotides to an end of the second polynucleotide of the two polynucleotides.

[0067] A "promoter" is a nucleotide sequence that controls or regulates the transcription of a nucleotide sequence (e.g., a coding sequence) that is operably associated with the promoter. The coding sequence controlled or regulated by a promoter may encode a polypeptide and/or a functional RNA. Typically, a "promoter" refers to a nucleotide sequence that contains a binding site for RNA polymerase II and directs the initiation of transcription. In general, promoters are found 5', or upstream, relative to the start of the coding region of the corresponding coding sequence. A promoter may comprise other elements that act as regulators of gene expression; e.g., a promoter region. These include a TATA box consensus sequence, and often a CAAT box consensus sequence (Breathnach and Chambon, (1981) *Annu. Rev. Biochem.* 50:349). In plants, the CAAT box may be substituted by the AGGA box (Messing et al., (1983) in Genetic Engineering of Plants, T. Kosuge, C. Meredith and A. Hollaender (eds.), Plenum Press, pp. 211-227). In some embodiments, a promoter region may comprise at least one intron (e.g., SEQ ID NO:36 or SEQ ID NO:37).

[0068] Promoters useful with this invention can include, for example, constitutive, inducible, temporally regulated, developmentally regulated, chemically regulated, tissue-preferred and/or tissue-specific promoters for use in the preparation of recombinant nucleic acid molecules, e.g., "synthetic nucleic acid constructs" or "protein-RNA complex." These various types of promoters are known in the art.

[0069] The choice of promoter may vary depending on the temporal and spatial requirements for expression, and also may vary based on the host cell to be transformed. Promoters for many different organisms are well known in the art. Based on the extensive knowledge present in the art, the appropriate promoter can be selected for the particular host organism of interest. Thus, for example, much is known about promoters upstream of highly constitutively expressed genes in model organisms and such knowledge can be readily accessed and implemented in other systems as appropriate.

[0070] In some embodiments, a promoter functional in a plant may be used with the constructs of this invention. Non-limiting examples of a promoter useful for driving expression in a plant include the promoter of the RubisCo small subunit gene 1 (PrbcS1), the promoter of the actin gene (Pactin), the promoter of the nitrate reductase gene (Pnr) and the promoter of duplicated carbonic anhydrase gene 1 (Pdca1) (See, Walker et al. *Plant Cell Rep.* 23:727-735 (2005); Li et al. *Gene* 403:132-142 (2007); Li et al. *Mol Biol. Rep.* 37:1143-1154 (2010)). PrbcS1 and Pactin are constitutive promoters and Pnr and Pdca1 are inducible promoters. Pnr is induced by nitrate and repressed by ammonium (Li et al. *Gene* 403:132-142 (2007)) and Pdca1 is induced by salt (Li et al. *Mol Biol. Rep.* 37:1143-1154 (2010)). In some embodiments, a promoter useful with this invention is RNA polymerase II (Pol II) promoter. In some embodiments, a U6 promoter or a 7SL promoter from *Zea mays* may be useful with constructs of this invention. In some embodiments, the U6c promoter and/or 7SL promoter from *Zea mays* may be useful for driving expression of a guide nucleic acid. In some embodiments, a U6c promoter, U6i promoter and/or 7SL promoter from *Glycine max* may be useful with constructs of this invention. In some embodiments, the U6c promoter, U6i promoter and/or 7SL promoter from *Glycine max* may be useful for driving expression of a guide nucleic acid.

[0071] Examples of constitutive promoters useful for plants include, but are not limited to, cestrum virus promoter (cmp) (U.S. Pat. No. 7,166,770), the rice actin 1 promoter (Wang et al. (1992) *Mol. Cell. Biol.* 12:3399-3406; as well as U.S. Pat. No. 5,641,876), CaMV 35S promoter (Odell et al. (1985) *Nature* 313:810-812), CaMV 19S promoter (Lawton et al. (1987) *Plant Mol. Biol.* 9:315-324), nos promoter (Ebert et al. (1987) *Proc. Natl. Acad. Sci USA* 84:5745-5749), Adh promoter (Walker et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:6624-6629), sucrose synthase promoter (Yang & Russell (1990) *Proc. Natl. Acad. Sci. USA* 87:4144-4148), and the ubiquitin promoter. The constitutive promoter derived from ubiquitin accumulates in many cell types. Ubiquitin promoters have been cloned from several plant species for use in transgenic plants, for example, sunflower (Binet et al., 1991. *Plant Science* 79: 87-94), maize (Christensen et al., 1989. *Plant Molec. Biol.* 12: 619-632), and *arabidopsis* (Norris et al. 1993. *Plant Molec. Biol.* 21:895-906). The maize ubiquitin promoter (UbiP) has been developed in transgenic monocot systems and its sequence and vectors constructed for monocot transformation are disclosed in the European patent publication EP0342926. The ubiquitin promoter is suitable for the expression of the nucleotide sequences of the invention in transgenic plants, especially monocotyledons. Further, the promoter expression cassettes described by McElroy et al. (*Mol. Gen. Genet.* 231: 150-160 (1991)) can be easily modified for the expression of the nucleotide sequences of the invention and are particularly suitable for use in monocotyledonous hosts.

[0072] In some embodiments, tissue specific/tissue preferred promoters can be used for expression of a heterologous polynucleotide in a plant cell. Tissue specific or preferred expression patterns include, but are not limited to, green tissue specific or preferred, root specific or preferred, stem specific or preferred, flower specific or preferred or pollen specific or preferred. Promoters suitable for expression in green tissue include many that regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. In one embodiment, a promoter useful with the invention is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, *Plant Molec. Biol.* 12:579-589 (1989)). Non-limiting examples of tissue-specific promoters include those associated with genes encoding the seed storage proteins (such as 0-conglycinin, cruciferin, napin and phaseolin), zein or oil body proteins (such as oleosin), or proteins involved in fatty acid biosynthesis (including acyl carrier protein, stearoyl-ACP desaturase and fatty acid desaturases (fad 2-1)), and other nucleic acids expressed during embryo development (such as Bce4, see, e.g., Kridl et al. (1991) *Seed Sci. Res.* 1:209-219; as well as EP Patent No. 255378). Tissue-specific or tissue-preferential promoters useful for the expression of the nucleotide sequences of the invention in plants, particularly maize, include but are not limited to those that direct expression in root, pith, leaf or pollen. Such promoters are disclosed, for example, in WO 93/07278, incorporated by reference herein for its disclosure of promoters. Other non-limiting examples of tissue specific or tissue preferred promoters useful with the invention the cotton rubisco promoter disclosed in U.S. Pat. No. 6,040,504; the rice sucrose synthase promoter disclosed in U.S. Pat. No. 5,604,121; the root specific promoter described by de Framond (FEBS 290:103-106 (1991); European patent EP 0452269 to Ciba-Geigy); the stem specific promoter described in U.S. Pat. No. 5,625,136 (to Ciba-Geigy) and which drives expression of the maize trpA gene; the cestrum yellow leaf curling virus promoter disclosed in WO 01/73087; and pollen specific or preferred promoters including, but not limited to, ProOsLPS10 and ProOsLPS11 from rice (Nguyen et al. *Plant Biotechnol. Reports* 9(5):297-306 (2015)), ZmSTK2_USP from maize (Wang et al. *Genome* 60(6):485-495 (2017)), LAT52 and LAT59 from tomato (Twell et al. *Development* 109(3):705-713 (1990)), Zm13 (U.S. Pat. No. 10,421,972), PLA$_2$-6 promoter from *arabidopsis* (U.S. Pat. No. 7,141,424), and/or the ZmC5 promoter from maize (International PCT Publication No. WO 1999/042587).

[0073] Additional examples of plant tissue-specific/tissue preferred promoters include, but are not limited to, the root hair-specific cis-elements (RHEs) (Kim et al. *The Plant Cell* 18:2958-2970 (2006)), the root-specific promoters RCc3 (Jeong et al. *Plant Physiol.* 153:185-197 (2010)) and RB7 (U.S. Pat. No. 5,459,252), the lectin promoter (Lindstrom et al. (1990) *Der. Genet.* 11:160-167; and Vodkin (1983) *Prog. Clin. Biol. Res.* 138:87-98), corn alcohol dehydrogenase 1 promoter (Dennis et al. (1984) *Nucleic Acids Res.* 12:3983-4000), S-adenosyl-L-methionine synthetase (SAMS) (Vander Mijnsbrugge et al. (1996) *Plant and Cell Physiology,* 37(8):1108-1115), corn light harvesting complex promoter (Bansal et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:3654-3658), corn heat shock protein promoter (O'Dell et al. (1985) *EMBO J.* 5:451-458; and Rochester et al. (1986) *EMBO J.* 5:451-458), pea small subunit RuBP carboxylase promoter (Cashmore, "Nuclear genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase" pp. 29-39 In: *Genetic Engineering of Plants* (Hollaender ed., Plenum Press 1983; and Poulsen et al. (1986) *Mol. Gen. Genet.* 205:193-200), Ti plasmid mannopine synthase promoter (Langridge et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3219-3223), Ti plasmid nopaline synthase promoter (Langridge et al. (1989), supra), *petunia* chalcone isomerase promoter (van Tunen et al. (1988) *EMBO J.* 7:1257-1263), bean glycine rich protein 1 promoter (Keller et al. (1989) *Genes Dev.* 3:1639-1646), truncated CaMV 35S promoter (O'Dell et al. (1985) *Nature* 313:810-812), potato patatin promoter (Wenzler et al. (1989) *Plant Mol. Biol.* 13:347-354), root cell promoter (Yamamoto et al. (1990) *Nucleic Acids Res.* 18:7449), maize zein promoter (Kriz et al. (1987) *Mol. Gen. Genet.* 207:90-98; Langridge et al. (1983) *Cell* 34:1015-1022; Reina et al. (1990) *Nucleic Acids Res.* 18:6425; Reina et al. (1990) *Nucleic Acids Res.* 18:7449; and Wandelt et al. (1989) *Nucleic Acids Res.* 17:2354), globulin-1 promoter

(Belanger et al. (1991) *Genetics* 129:863-872), α-tubulin cab promoter (Sullivan et al. (1989) *Mol. Gen. Genet.* 215:431-440), PEPCase promoter (Hudspeth & Grula (1989) *Plant Mol. Biol.* 12:579-589), R gene complex-associated promoters (Chandler et al. (1989) *Plant Cell* 1:1175-1183), and chalcone synthase promoters (Franken et al. (1991) *EMBO J.* 10:2605-2612).

[0074] Useful for seed-specific expression is the pea vicilin promoter (Czako et al. (1992) *Mol. Gen. Genet.* 235:33-40; as well as the seed-specific promoters disclosed in U.S. Pat. No. 5,625,136. Useful promoters for expression in mature leaves are those that are switched at the onset of senescence, such as the SAG promoter from *Arabidopsis* (Gan et al. (1995) *Science* 270:1986-1988).

[0075] In addition, promoters functional in chloroplasts can be used. Non-limiting examples of such promoters include the bacteriophage T3 gene 9 5' UTR and other promoters disclosed in U.S. Pat. No. 7,579,516. Other promoters useful with the invention include but are not limited to the S-E9 small subunit RuBP carboxylase promoter and the Kunitz trypsin inhibitor gene promoter (Kti3).

[0076] Additional regulatory elements useful with this invention include, but are not limited to, introns, enhancers, termination sequences and/or 5' and 3' untranslated regions.

[0077] An intron useful with this invention can be an intron identified in and isolated from a plant and then inserted into an expression cassette to be used in transformation of a plant. As would be understood by those of skill in the art, introns can comprise the sequences required for self-excision and are incorporated into nucleic acid constructs/expression cassettes in frame. An intron can be used either as a spacer to separate multiple protein-coding sequences in one nucleic acid construct, or an intron can be used inside one protein-coding sequence to, for example, stabilize the mRNA. If they are used within a protein-coding sequence, they are inserted "in-frame" with the excision sites included. Introns may also be associated with promoters to improve or modify expression. As an example, a promoter/intron combination useful with this invention includes but is not limited to that of the maize Ubi1 promoter and intron.

[0078] Non-limiting examples of introns useful with the present invention include introns from the ADHI gene (e.g., Adh1-S introns 1, 2 and 6), the ubiquitin gene (Ubi1), the RuBisCO small subunit (rbcS) gene, the RuBisCO large subunit (rbcL) gene, the actin gene (e.g., actin-1 intron), the pyruvate dehydrogenase kinase gene (pdk), the nitrate reductase gene (nr), the duplicated carbonic anhydrase gene 1 (Tdca1), the psbA gene, the atpA gene, or any combination thereof.

[0079] An "editing system" as used herein refers to any site-specific (e.g., sequence-specific) nucleic acid editing system now known or later developed, which system can introduce a modification (e.g., a mutation) in a nucleic acid in target specific manner. For example, an editing system (e.g., a site- and/or sequence-specific editing system) can include, but is not limited to, a CRISPR-Cas editing system, a meganuclease editing system, a zinc finger nuclease (ZFN) editing system, a transcription activator-like effector nuclease (TALEN) editing system, a base editing system and/or a prime editing system, each of which may comprise one or more polypeptide(s) and/or one or more polynucleotide(s) that when present and/or expressed together (e.g., as a system) in a composition and/or cell can modify (e.g., mutate) a target nucleic acid in a sequence specific manner. In some embodiments, an editing system (e.g., a site- and/or sequence-specific editing system) can comprise one or more polynucleotide(s) and/or one or more polypeptide(s), including but not limited to a nucleic acid binding polypeptide (e.g., a DNA binding domain), a nuclease, another polypeptide, and/or a polynucleotide. In some embodiments, a CRISPR-Cas editing system is provided, wherein a fusion protein of the present invention is used to provide a Cas12a protein of the CRISPR-Cas editing system. An editing system of the present invention may modify a target nucleic acid that is present in a cell or outside a cell (e.g., a method of the present invention may be carried out in vitro, ex vivo, and/or in vivo).

[0080] In some embodiments, an editing system comprises one or more sequence-specific nucleic acid binding polypeptide(s) (e.g., a DNA binding domain) that can be from, for example, a polynucleotide-guided endonuclease, a CRISPR-Cas endonuclease (e.g., CRISPR-Cas effector protein), a zinc finger nuclease, a transcription activator-like effector nuclease (TALEN) and/or an Argonaute protein. In some embodiments, an editing system comprises one or more cleavage polypeptide(s) (e.g., nucleases) including, but not limited to, an endonuclease (e.g., Fok1), a polynucleotide-guided endonuclease, a CRISPR-Cas endonuclease (e.g., CRISPR-Cas effector protein), a zinc finger nuclease, and/or a transcription activator-like effector nuclease (TALEN).

[0081] A "nucleic acid binding polypeptide" as used herein refers to a polypeptide or domain that binds and/or is capable of binding a nucleic acid (e.g., a target nucleic acid). A DNA binding domain or DNA binding polypeptide is an exemplary nucleic acid binding polypeptide and may be a site- and/or sequence-specific nucleic acid binding domain. In some embodiments, a nucleic acid binding polypeptide may be a sequence-specific nucleic acid binding polypeptide such as, but not limited to, a sequence-specific binding domain from, for example, a polynucleotide-guided endonuclease, a CRISPR-Cas effector protein (e.g., a CRISPR-Cas endonuclease), a zinc finger nuclease, a transcription activator-like effector nuclease (TALEN) and/or an Argonaute protein. In some embodiments, a nucleic acid binding polypeptide comprises a cleavage domain (e.g., a nuclease domain) such as, but not limited to, an endonuclease (e.g., Fok1), a polynucleotide-guided endonuclease, a CRISPR-Cas endonuclease, a zinc finger nuclease, and/or a transcription activator-like effector nuclease (TALEN). In some embodiments, the nucleic acid binding polypeptide associates with and/or is capable of associating with (e.g., forms a complex with) one or more nucleic acid molecule(s) (e.g., forms a complex with a guide nucleic acid as described herein), which may direct and/or guide the nucleic acid binding polypeptide to a specific target nucleotide sequence (e.g., a gene locus of a genome) that is complementary to the one or more nucleic acid molecule(s) (or a portion or region thereof), thereby causing the nucleic acid binding polypeptide to bind to the nucleotide sequence at the specific target site. In some embodiments, the nucleic acid binding polypeptide is a CRISPR-Cas effector protein as described herein.

[0082] In some embodiments, an editing system comprises or is a ribonucleoprotein such as an assembled ribonucleoprotein complex (e.g., a ribonucleoprotein that comprises a CRISPR-Cas effector protein, a guide nucleic acid, and optionally a deaminase). In some embodiments, a ribonucleoprotein of an editing system may be assembled together (e.g., a pre-assembled ribonucleoprotein including a CRISPR-Cas effector protein, a guide nucleic acid, and optionally a deaminase) such as when contacted to a target nucleic acid or when introduced into a cell (e.g., a mammalian cell or a plant cell) (e.g., at the time of contacting the components of the ribonucleoprotein to a target nucleic acid and/or at the time of introducing the components of the ribonucleoprotein into a cell). In some embodiments, a ribonucleoprotein of an editing system may assemble into a complex (e.g., a non-covalently bound complex) while a portion of the ribonucleoprotein is contacting a target nucleic acid and/or may assemble after and/or during introduction into a plant cell. In some embodiments, an editing system may be assembled (e.g., into a non-covalently bound complex) when introduced into a plant cell. In some embodiments, a ribonucleoprotein may comprise a fusion protein of the present invention, a guide nucleic acid, and optionally a deaminase. In some embodiments, a ribonucleoprotein of an editing system may be contacted to a target nucleic acid and/or may be introduced into a plant cell. In some embodiments, an editing system may be assembled (e.g., into a non-covalently bound complex) when introduced into a plant cell. In some embodiments, a ribonucleoprotein may comprise a protein of the present invention (e.g., a protein prepared using a composition and/or method of the present invention), a guide nucleic acid, and optionally a deaminase and/or reverse transcriptase. In some embodiments, a protein of the present invention comprises a CRISPR-Cas effector protein and the protein is used in place of (e.g., substituted for) a CRISPR-Cas effector protein (e.g., in a composition, complex, kit, method, and/or system such as an editing system described herein) and/or functions as a CRISPR-Cas effector protein, templated editor, and/or base editor, optionally in a composition, complex, ribonucleoprotein, kit, method, system, and/or editing system of the present invention.

[0083] The terms “transgene” or “transgenic” as used herein refer to at least one nucleic acid sequence that is taken from the genome of one organism or produced synthetically, and which is then introduced into a host cell (e.g., a plant cell) or organism or tissue of interest and which is subsequently integrated into the host’s genome by means of “stable” transformation or transfection approaches. In contrast, the term “transient” transformation or transfection or introduction refers to a way of introducing molecular tools including at least one nucleic acid (DNA, RNA, single-stranded or double-stranded or a mixture thereof) and/or at least one amino acid sequence, optionally comprising suitable chemical or biological agents, to achieve a transfer into at least one compartment of interest of a cell, including, but not restricted to, the cytoplasm, an organelle, including the nucleus, a mitochondrion, a vacuole, a chloroplast, or into a membrane, resulting in transcription and/or translation and/or association and/or activity of the at least one molecule introduced without achieving a stable integration or incorporation into the genome and thus without inheritance of the respective at least one molecule introduced into the genome of a cell. The term “transgene-free” refers to a condition in which a transgene is not present or found in the genome of a host cell or tissue or organism of interest.

[0084] In some embodiments, a polynucleotide and/or a nucleic acid construct of the invention can be an “expression cassette” or can be comprised within an expression cassette. As used herein, “expression cassette” means a recombinant nucleic acid molecule comprising, for example, a nucleic acid construct of the invention (e.g., a polynucleotide encoding a fusion protein of the present invention, a polynucleotide encoding a cytosine deaminase, a polynucleotide encoding an adenine deaminase, a polynucleotide encoding a deaminase fusion protein, a polynucleotide encoding a peptide tag, a polynucleotide encoding an affinity polypeptide, a polynucleotide encoding a glycosylase, and/or a polynucleotide comprising a guide nucleic acid), wherein the nucleic acid construct is operably associated with at least a control sequence (e.g., a promoter). Thus, some embodiments of the invention provide expression cassettes designed to express, for example, a nucleic acid construct of the invention. When an expression cassette comprises more than one polynucleotide, the polynucleotides may be operably linked to a single promoter that drives expression of all of the polynucleotides or the polynucleotides may be operably linked to one or more separate promoters (e.g., three polynucleotides may be driven by one, two or three promoters in any combination). Thus, for example, a polynucleotide encoding a fusion protein, a polynucleotide encoding a deaminase (e.g., an adenine deaminase), and a polynucleotide comprising a guide nucleic acid comprised in an expression cassette may each be operably associated with a single promoter or one or more of the polynucleotide(s) may be operably associated with separate promoters (e.g., two or three promoters) in any combination, which may be the same or different from each other.

[0085] In some embodiments, an expression cassette comprising the polynucleotides/nucleic acid constructs of the invention may be optimized for expression in an organism (e.g., an animal, a plant, a bacterium and the like).

[0086] An expression cassette comprising a nucleic acid construct of the invention may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components (e.g., a promoter from the host organism operably linked to a polynucleotide of interest to be expressed in the host organism, wherein the polynucleotide of interest is from a different organism than the host or is not normally found in association with that promoter). An expression cassette may also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression.

[0087] An expression cassette can optionally include a transcriptional and/or translational termination region (i.e., termination region) and/or an enhancer region that is functional in the selected host cell. A variety of transcriptional terminators and enhancers are known in the art and are available for use in expression cassettes. Transcriptional terminators are responsible for the termination of transcription and correct mRNA polyadenylation. A termination region and/or the enhancer region may be native to the transcriptional initiation region, may be native to a gene encoding a CRISPR-Cas effector protein or a gene encoding a deaminase, may be native to a host cell, or may be native to another source (e.g., foreign or heterologous to the promoter, to a gene encoding the CRISPR-Cas effector protein or a gene encoding the deaminase, to a host cell, or any combination thereof).

[0088] An expression cassette of the invention also can include a polynucleotide encoding a selectable marker, which can be used to select a transformed host cell. As used herein, "selectable marker" means a polynucleotide sequence that when expressed imparts a distinct phenotype to the host cell expressing the marker and thus allows such transformed cells to be distinguished from those that do not have the marker. Such a polynucleotide sequence may encode either a selectable or screenable marker, depending on whether the marker confers a trait that can be selected for by chemical means, such as by using a selective agent (e.g., an antibiotic and the like), or on whether the marker is simply a trait that one can identify through observation or testing, such as by screening (e.g., fluorescence). Many examples of suitable selectable markers are known in the art and can be used in the expression cassettes described herein.

[0089] The expression cassettes, the nucleic acid molecules/constructs and polynucleotide sequences described herein can be used in connection with vectors. The term "vector" refers to a composition for transferring, delivering or introducing a nucleic acid (or nucleic acids) into a cell. A vector may comprise a nucleic acid construct comprising one or more nucleotide sequence(s) to be transferred, delivered or introduced into a cell. Vectors for use in transformation of host organisms are well known in the art. Non-limiting examples of general classes of vectors include viral vectors (e.g., Adeno-associated virus (AAV) vectors), plasmid vectors, phage vectors, phagemid vectors, cosmid vectors, fosmid vectors, bacteriophages, artificial chromosomes, minicircles, or *Agrobacterium* binary vectors in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable. In some embodiments, a viral vector can include, but is not limited, to a retroviral, lentiviral, adenoviral, adeno-associated, or herpes simplex viral vector. A vector as defined herein can transform a prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g., autonomous replicating plasmid with an origin of replication). Additionally, included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g., higher plant, mammalian, yeast or fungal cells). In some embodiments, the nucleic acid in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of genomic DNA, this may contain its own promoter and/or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter and/or other regulatory elements for expression in the host cell. Accordingly, a nucleic acid construct of this invention and/or expression cassettes comprising the same may be comprised in vectors as described herein and as known in the art.

[0090] As used herein, "contact," "contacting," "contacted," and grammatical variations thereof, refer to placing the components of a desired reaction together under conditions suitable for carrying out the desired reaction (e.g., transformation, transcriptional control, genome editing, nicking, and/or cleavage). Thus, for example, a target nucleic acid may be contacted with a nucleic acid construct of the invention encoding, for example, a nucleic acid binding polypeptide (e.g., a DNA binding polypeptide such as a sequence-specific DNA binding protein (e.g., a polynucleotide-guided endonuclease, a CRISPR-Cas effector protein (e.g., a CRISPR-Cas endonuclease), a zinc finger nuclease, a transcription activator-like effector nuclease (TALEN) and/or an Argonaute protein)), a guide nucleic acid, and optionally a cytosine deaminase and/or adenine deaminase under conditions whereby the nucleic acid binding polypeptide (e.g., a CRISPR-Cas effector protein) is expressed, and the nucleic acid binding polypeptide forms a complex with the guide nucleic acid, the complex hybridizes to the target nucleic acid, and optionally the cytosine deaminase and/or adenine deaminase is/are recruited to the nucleic acid binding polypeptide (and thus, to the target nucleic acid) or the cytosine deaminase and/or adenine deaminase are fused to the nucleic acid binding polypeptide, thereby modifying the target nucleic acid. In some embodiments, the cytosine deaminase and/or adenine deaminase and the nucleic acid binding polypeptide localize at the target nucleic acid, optionally through covalent and/or non-covalent interactions.

[0091] In some embodiments, a target nucleic acid may be contacted with a nucleic acid construct of the invention encoding a fusion protein of the present invention, a guide nucleic acid, and optionally a cytosine deaminase and/or adenine deaminase under conditions whereby the fusion protein is produced, or a target nucleic acid may be contacted with a fusion protein of the present invention, a guide nucleic acid, and optionally a cytosine deaminase and/or adenine deaminase. The fusion protein can form a complex with the guide nucleic acid, and the complex can hybridize to the target nucleic acid, and optionally the cytosine deaminase and/or adenine deaminase is/are recruited to the fusion protein (and thus, to the target nucleic acid) or the cytosine deaminase and/or adenine deaminase are fused to the fusion protein, thereby modifying the target nucleic acid. The cytosine deaminase and/or adenine deaminase and the fusion protein may localize at the target nucleic acid, optionally through covalent and/or non-covalent interactions.

[0092] As used herein, "modifying" or "modification" in reference to a target nucleic acid includes editing (e.g., mutating), covalent modification, exchanging/substituting nucleic acids/nucleotide bases, deleting, cleaving, and/or nicking of a target nucleic acid to thereby provide a modified nucleic acid and/or altering transcriptional control of a target nucleic acid to thereby provide a modified nucleic acid. In some embodiments, a modification may include an insertion and/or deletion of any size and/or a single base change (SNP) of any type. In some embodiments, a modification comprises a SNP. In some embodiments, a modification comprises exchanging and/or substituting one or more (e.g., 1, 2, 3, 4, 5, or more) nucleotides. In some embodiments, an insertion or deletion may be about 1 base to about 30,000 bases or more in length (e.g., about 1, 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, 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, 10,000, 10,500, 11,000, 11,500, 12,000, 12,500, 13,000, 13,500, 14,000, 14,500, 15,000, 15,500, 16,000, 16,500, 17,000, 17,500, 18,000, 18,500, 19,000, 19,500, 20,000, 20,500, 21,000, 21,500, 22,000, 22,500, 23,000, 23,500, 24,000, 24,500, 25,000, 25,500, 26,000, 26,500, 27,000, 27,500, 28,000, 28,500, 29,000, 29,500, 30,000 bases in length or more, or any value or range therein). Thus, in some embodiments, an insertion or deletion may be about 1, 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, 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 to about 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 bases in length, or any range or value therein; about 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300 bases to about 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000 bases or more in length, or any value or range therein; about 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000 bases to about 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, or 10,000 bases or more in length, or any value or range therein; or about 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, or 700 bases to about 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, or 5000 bases or more in length, or any value or range therein. In some embodiments, an insertion or deletion may be about 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000, 8500, 9000, 9500, or 10,000 bases to about 10,500, 11,000, 11,500, 12,000, 12,500, 13,000, 13,500, 14,000, 14,500, 15,000, 15,500, 16,000, 16,500, 17,000, 17,500, 18,000, 18,500, 19,000, 19,500, 20,000, 20,500, 21,000, 21,500, 22,000, 22,500, 23,000, 23,500, 24,000, 24,500, 25,000, 25,500, 26,000, 26,500, 27,000, 27,500, 28,000, 28,500, 29,000, 29,500, or 30,000 bases or more in length, or any value or range therein.

[0093] "Recruit," "recruiting" or "recruitment" as used herein refer to attracting one or more polypeptide(s) or polynucleotide(s) to another polypeptide or polynucleotide (e.g., to a particular location in a genome) using protein-protein interactions, nucleic acid protein interactions (e.g., RNA-protein interactions), and/or chemical interactions. Protein-protein interactions can include, but are not limited to, peptide tags (epitopes, multimerized epitopes) and corresponding affinity polypeptides, RNA recruiting motifs and corresponding affinity polypeptides, and/or chemical interactions. Example chemical interactions that may be useful with polypeptides and polynucleotides for the purpose of recruitment can include, but are not limited to, rapamycin-inducible dimerization of FRB-FKBP; Biotin-streptavidin interaction; SNAP tag (Hussain et al. *Curr Pharm Des.* 19(30):5437-42 (2013)); Halo tag (Los et al. *ACS Chem Biol.* 3(6):373-82 (2008)); CLIP tag (Gautier et al. *Chemistry & Biology* 15:128-136 (2008)); DmrA-DmrC heterodimer induced by a compound (Tak et al. *Nat Methods* 14(12):1163-1166 (2017)); Bifunctional ligand approaches (fuse two protein-binding chemicals together) (VoB et al. *Curr Opin Chemical Biology* 28:194-201 (2015)) (e.g. dihyrofolate reductase (DHFR) (Kopyteck et al. Cell Cehm Biol 7(5):313-321 (2000)).

[0094] "Introducing," "introduce," "introduced" (and grammatical variations thereof) in the context of a polynucleotide of interest or editing system means presenting a nucleotide sequence of interest (e.g., polynucleotide, a nucleic acid construct, and/or a guide nucleic acid) and/or editing system (e.g., a polynucleotide, polypeptide, and/or ribonucleoprotein) to a host organism or cell of said organism (e.g., host cell; e.g., a plant cell) in such a manner that the nucleotide sequence and/or editing system gains access to the interior of a cell. Thus, for example, a nucleic acid construct of the invention encoding a fusion protein of the present invention, a guide nucleic acid, and/or a cytosine deaminase and/or adenine deaminase may be introduced into a cell of an organism, thereby transforming the cell with the fusion protein, guide nucleic acid, and/or cytosine deaminase and/or adenine deaminase. In some embodiments, a fusion protein of the present invention and/or a guide nucleic acid may be introduced into a cell of an organism, optionally wherein the fusion protein and guide nucleic acid may be comprised in a complex (e.g., a ribonucleoprotein). In some embodiments, the organism is a eukaryote (e.g., a mammal such as a human).

[0095] The term "transformation" as used herein refers to the introduction of a nucleic acid, polypeptide, and/or ribonucleoprotein (e.g., heterologous nucleic acid, polypeptide, and/or ribonucleoprotein) into a cell. Transformation of a cell may be stable or transient. Thus, in some embodiments, a host cell or host organism may be stably transformed with a polynucleotide/nucleic acid molecule of the invention. In some embodiments, a host cell or host organism may be transiently transformed with a nucleic acid construct, a polypeptide, and/or a ribonucleoprotein of the invention.

[0096] "Transient transformation" in the context of a polynucleotide, polypeptide, and/or ribonucleoprotein means that a polynucleotide, polypeptide, and/or ribonucleoprotein is introduced into the cell and does not integrate into the genome of the cell.

[0097] By "stably introducing" or "stably introduced" in the context of a polynucleotide introduced into a cell is intended that the introduced polynucleotide is stably incorporated into the genome of the cell, and thus the cell is stably transformed with the polynucleotide.

[0098] "Stable transformation" or "stably transformed" as used herein means that a nucleic acid molecule is introduced into a cell and integrates into the genome of the cell. As such, the integrated nucleic acid molecule is capable of being inherited by the progeny thereof, more particularly, by the progeny of multiple successive generations. "Genome" as used herein includes the nuclear and the plastid genome, and therefore includes integration of the nucleic acid into, for example, the chloroplast or mitochondrial genome. Stable transformation as used herein can also refer to a transgene that is maintained extrachromasomally, for example, as a minichromosome or a plasmid.

[0099] Transient transformation may be detected by, for example, an enzyme-linked immunosorbent assay (ELISA) or Western blot, which can detect the presence of a peptide or polypeptide encoded by one or more transgene introduced into an organism. Stable transformation of a cell can be detected by, for example, a Southern blot hybridization assay of genomic DNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a transgene introduced into an organism (e.g., a mammal, plant, etc.). Stable transformation of a cell can be detected by, for example, a Northern blot hybridization assay of RNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a transgene introduced into a host organism. Stable transformation of a cell can also be detected by, e.g., a polymerase chain reaction (PCR) or other amplification reactions as are well known in the art, employing specific primer sequences that hybridize with target sequence(s) of a transgene, resulting in amplification of the transgene sequence, which can be detected according to standard methods. Transformation can also be detected by direct sequencing and/or hybridization protocols well known in the art.

[0100] Accordingly, in some embodiments, nucleotide sequences, polynucleotides, nucleic acid constructs, and/or expression cassettes of the invention may be expressed transiently and/or they can be stably incorporated into the genome of the host organism. Thus, in some embodiments, a nucleic acid construct of the invention may be transiently introduced into a cell with a guide nucleic acid and as such, no DNA maintained in the cell.

[0101] A nucleic acid construct, polypeptide, and/or ribonucleoprotein of the invention can be introduced into a cell by any method known to those of skill in the art. In some embodiments, transformation methods include, but are not limited to, transformation via bacterial-mediated nucleic acid delivery (e.g., via Agrobacteria), viral-mediated nucleic acid delivery, silicon carbide and/or nucleic acid whisker-mediated nucleic acid delivery, liposome mediated nucleic acid delivery, microinjection, microparticle bombardment, calcium-phosphate-mediated transformation, cyclodextrin-mediated transformation, electroporation, nanoparticle-mediated transformation, sonication, infiltration, PEG-mediated nucleic acid uptake, as well as any other electrical, chemical, physical (mechanical) and/or biological mechanism that results in the introduction of nucleic acid into the cell (e.g., a plant cell or an animal cell), including any combination thereof. In some embodiments of the invention, transformation of a cell comprises nuclear transformation. In some embodiments, transformation of a cell comprises plastid transformation (e.g., chloroplast transformation). In some embodiments, a recombinant nucleic acid construct of the invention can be introduced into a cell via conventional breeding techniques.

[0102] Procedures for transforming both eukaryotic and prokaryotic organisms are well known and routine in the art and are described throughout the literature (See, for example, Jiang et al. 2013. *Nat. Biotechnol.* 31:233-239; Ran et al. *Nature Protocols* 8:2281-2308 (2013)). General guides to various plant transformation methods known in the art include Miki et al. ("Procedures for Introducing Foreign DNA into Plants" in Methods in Plant Molecular Biology and Biotechnology, Glick, B. R. and Thompson, J. E., Eds. (CRC Press, Inc., Boca Raton, 1993), pages 67-88) and Rakowoczy-Trojanowska (Cell. Mol. Biol. Lett. 7:849-858 (2002)).

[0103] A nucleotide sequence, polypeptide, and/or ribonucleoprotein therefore can be introduced into a host organism or its cell in any number of ways that are well known in the art. The methods of the invention do not depend on a particular method for introducing one or more nucleotide sequence(s), polypeptide(s), and/or ribonucleoprotein(s) into the organism, only that they gain access to the interior of at least one cell of the organism. Where more than one nucleotide sequence, polypeptide, and/or ribonucleoprotein is to be introduced, they can be assembled as part of a single nucleic acid construct, or as separate nucleic acid constructs, and can be located on the same or different nucleic acid constructs. Accordingly, a nucleotide sequence, polypeptide, and/or ribonucleoprotein can be introduced into the cell of interest in a single transformation event, and/or in separate transformation events, or, alternatively, where relevant, a nucleotide sequence can be incorporated into a plant, for example, as part of a breeding protocol. In some embodiments, the cell is a eukaryotic cell (e.g., a plant cell or a mammalian such as a human cell).

[0104] In some embodiments, a nucleic acid construct of the invention (e.g., a polynucleotide encoding a fusion protein of the present invention, a polynucleotide encoding a deaminase, and/or a guide nucleic acid and/or expression cassettes and/or vectors comprising the same) may be operably linked to at least one regulatory sequence, optionally, wherein the at least one regulatory sequence may be codon optimized for expression in a plant. In some embodiments, the at least one regulatory sequence may be, for example, a promoter, an operon, a terminator, or an enhancer. In some embodiments, the at least one regulatory sequence may be a promoter. In some embodiments, the regulatory sequence may be an intron. In some embodiments, the at least one regulatory sequence may be, for example, a promoter operably associated with an intron or a promoter region comprising an intron. In some embodiments, the at least one regulatory sequence may be, for example a ubiquitin promoter and its associated intron (e.g., *Medicago truncatula* and/or *Zea mays* and their associated introns). In some embodiments, the at least one regulatory sequence may be a terminator nucleotide sequence and/or an enhancer nucleotide sequence.

[0105] In some embodiments, a nucleic acid construct of the invention may be operably associated with a promoter region, wherein the promoter region comprises an intron, optionally wherein the promoter region may be a ubiquitin promoter and intron (e.g., a *Medicago* or a maize ubiquitin promoter and intron, e.g., SEQ ID NO:36 or SEQ ID NO:37). In some embodiments, the nucleic acid construct of the invention that is operably associated with a promoter region comprising an intron may be codon optimized for expression in a plant.

[0106] In some embodiments, a nucleic acid construct of the invention may encode one or more (e.g., 1, 2, 3, 4, or more) polypeptide(s) of interest. The one or more polypeptides of interest may be codon optimized for expression in a eukaryote (e.g., a human or a plant). In some embodiments, a fusion protein may comprise one or more (e.g., 1, 2, 3, 4, or more) polypeptide(s) of interest.

[0107] A polypeptide of interest useful with this invention can include, but is not limited to, a polypeptide or protein domain having deaminase activity, nickase activity, recombinase activity, transposase activity, methylase activity, glycosylase (DNA glycosylase) activity, glycosylase inhibitor activity (e.g., uracil-DNA glycosylase inhibitor (UGI)), a reverse transcriptase, a peptide tag (e.g., a GCN4 peptide tag), demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, restriction endonuclease activity (e.g., Fok1), nucleic acid binding activity, methyltransferase activity, DNA repair activity, DNA damage activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, polymerase activity (e.g., DNA polymerase activity), ligase activity, helicase activity, a nuclear localization sequence or activity, T-DNA processing and/or transfer (e.g., VirD2), an affinity polypeptide, a peptide tag, and/or photolyase activity. In some embodiments, the polypeptide of interest is a Fok1 nuclease, or a uracil-DNA glycosylase inhibitor. When encoded in a nucleic acid (polynucleotide, expression cassette, and/or vector) the encoded polypeptide or protein domain may be codon optimized for expression in an organism. In some embodiments, a polypeptide of interest may be linked to a fusion protein of the present invention or to a CRISPR-Cas effector protein domain to provide a CRISPR-Cas fusion protein. In some embodiments, a CRISPR-Cas fusion protein that comprises a CRISPR-Cas effector protein domain linked to a recruiting motif (e.g., a peptide tag) may also be linked to a polypeptide of interest (e.g., a CRISPR-Cas effector protein domain may be, for example, linked to both a recruiting motif (e.g., a peptide tag or an affinity polypeptide) and, for example, a polypeptide of interest.

[0108] In some embodiments, an editing system of the present invention comprises a CRISPR-Cas effector protein. As used herein, a "CRISPR-Cas effector protein" is a protein or polypeptide that cleaves, cuts, or nicks a nucleic acid; binds a nucleic acid (e.g., a target nucleic acid and/or a guide nucleic acid); and/or that identifies, recognizes, or binds a guide nucleic acid as defined herein. In some embodiments, a CRISPR-Cas effector protein may be an enzyme (e.g., a nuclease, endonuclease, nickase, etc.) and/or may function as an enzyme. In some embodiments, a CRISPR-Cas effector protein refers to a CRISPR-Cas nuclease. In some embodiments, a CRISPR-Cas effector protein comprises nuclease activity and/or nickase activity, comprises a nuclease domain whose nuclease activity and/or nickase activity has been reduced or eliminated, comprises single stranded DNA cleavage activity (ss DNAse activity) or which has ss DNAse activity that has been reduced or eliminated, and/or comprises self-processing RNAse activity or which has self-processing RNAse activity that has been reduced or eliminated. A CRISPR-Cas effector protein may bind to a target nucleic acid. A CRISPR-Cas effector protein may be a Type I, II, III, IV, V, or VI CRISPR-Cas effector protein. In some embodiments, a CRISPR-Cas effector protein may be from a Type I CRISPR-Cas system, a Type II CRISPR-Cas system, a Type III CRISPR-Cas system, a Type IV CRISPR-Cas system, Type V CRISPR-Cas system, or a Type VI CRISPR-Cas system. In some embodiments, a CRISPR-Cas effector protein of the invention may be from a Type II CRISPR-Cas system or a Type V CRISPR-Cas system. In some embodiments, a CRISPR-Cas effector protein may be a Type II CRISPR-Cas effector protein, for example, a Cas9 effector protein. In some embodiments, a CRISPR-Cas effector protein may be Type V CRISPR-Cas effector protein, for example, a Cas12 effector protein. In some embodiments, a CRISPR-Cas effector protein may be Cas12a and optionally may have an amino acid sequence of any one of SEQ ID NOs:38-60 or 192-195 and/or a nucleotide sequence of any one of SEQ ID NOs:61-63. In some embodiments, a CRISPR-Cas effector protein may be an active Cas12a and optionally may have an amino acid sequence of SEQ ID NO:46 or 55. In some embodiments, a CRISPR-Cas effector protein may be an inactive (i.e., dead) Cas12a and optionally may have an amino acid sequence of SEQ ID NO:38.

[0109] Exemplary CRISPR-Cas effector proteins include, but are not limited to, a Cas9, C2c1, C2c3, Cas12a (also referred to as Cpf1), Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Cas1, Cas1B, Cas2, Cas3, Cas3', Cas3", Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4 (dinG), and/or Csf5 nuclease, optionally wherein the CRISPR-Cas effector protein may be a Cas9, Cas12a (Cpf1), Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, and/or Cas14c effector protein.

[0110] In some embodiments, a CRISPR-Cas effector protein useful with the invention may comprise a mutation in its nuclease active site and/or nuclease domain (e.g., a RuvC, HNH, e.g., a RuvC site of a Cas12a nuclease domain; e.g., a RuvC site and/or HNH site of a Cas9 nuclease domain). A CRISPR-Cas effector protein having a mutation in its nuclease active site and/or nuclease domain, and therefore, no longer comprising nuclease activity, is commonly referred to as "inactive" or "dead," e.g., dCas12a. In some embodiments, a CRISPR-Cas effector protein having a mutation in its nuclease active site and/or nuclease domain may have impaired activity or reduced activity (e.g., nickase activity) as compared to the same CRISPR-Cas effector protein without the mutation.

[0111] A Type V CRISPR-Cas effector protein useful with embodiments of the invention may be a Cas12a. The CRISPR-Cas effector protein may be a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas nuclease. Cas12a differs in several respects from the more well-known Type II CRISPR Cas9 nuclease.

For example, Cas9 recognizes a G-rich protospacer-adjacent motif (PAM) that is 3' to its guide RNA (gRNA, sgRNA, crRNA, crDNA, CRISPR array) binding site (protospacer, target nucleic acid, target DNA) (3'-NGG), while Cas12a recognizes a T-rich PAM that is located 5' to the target nucleic acid (5'-TTN, 5'-TTTN. In fact, the orientations in which Cas9 and Cas12a bind their guide RNAs are very nearly reversed in relation to their N and C termini. Furthermore, Cas12a enzymes use a single guide RNA (gRNA, CRISPR array, crRNA) rather than the dual guide RNA (sgRNA (e.g., crRNA and tracrRNA)) found in natural Cas9 systems, and Cas12a processes its own gRNAs. Additionally, Cas12a nuclease activity produces staggered DNA double stranded breaks instead of blunt ends produced by Cas9 nuclease activity, and Cas12a relies on a single RuvC domain to cleave both DNA strands, whereas Cas9 utilizes an HNH domain and a RuvC domain for cleavage.

[0112] A CRISPR Cas12a effector protein useful with this invention may be any known or later identified Cas12a (previously known as Cpf1) (see, e.g., U.S. Pat. No. 9,790,490, which is incorporated by reference for its disclosures of Cpf1 (Cas12a) sequences). The term "Cas12a" refers to an RNA-guided protein that can have nuclease activity, the protein comprising a guide nucleic acid binding domain and an active, inactive, or partially active DNA cleavage domain, thereby the RNA-guided nuclease activity of the Cas12a may be active, inactive or partially active, respectively. In some embodiments, a Cas12a useful with the invention may comprise a mutation in the nuclease active site (e.g., RuvC site of the Cas12a domain). A Cas12a having a mutation in its nuclease domain and/or nuclease active site, and therefore, no longer comprising nuclease activity, is commonly referred to as deadCas12a (e.g., dCas12a). In some embodiments, a Cas12a having a mutation in its nuclease domain and/or nuclease active site may have impaired activity, e.g., may have reduced nickase activity.

[0113] In some embodiments, a CRISPR-Cas effector protein (e.g., Cas12a) may be optimized for expression in an organism, for example, in an animal (e.g., a mammal such as a human), a plant, a fungus, an archaeon, or a bacterium. In some embodiments, a CRISPR-Cas effector protein (e.g., Cas12a) may be optimized for expression in a plant.

[0114] Any deaminase domain/polypeptide useful for base editing may be used with this invention. A "cytosine deaminase" and "cytidine deaminase" as used herein refer to a polypeptide or domain thereof that catalyzes or is capable of catalyzing cytosine deamination in that the polypeptide or domain catalyzes or is capable of catalyzing the removal of an amine group from a cytosine base. Thus, a cytosine deaminase may result in conversion of cytosine to a thymidine (through a uracil intermediate), causing a C to T conversion, or a G to A conversion in the complementary strand in the genome. Thus, in some embodiments, the cytosine deaminase encoded by the polynucleotide of the invention generates a C→T conversion in the sense (e.g., "+"; template) strand of the target nucleic acid or a G→A conversion in antisense (e.g., "−", complementary) strand of the target nucleic acid. In some embodiments, a cytosine deaminase encoded by a polynucleotide of the invention generates a C to T, G, or A conversion in the complementary strand in the genome.

[0115] A cytosine deaminase useful with this invention may be any known or later identified cytosine deaminase from any organism (see, e.g., U.S. Pat. No. 10,167,457 and Thuronyi et al. *Nat. Biotechnol.* 37:1070-1079 (2019), each of which is incorporated by reference herein for its disclosure of cytosine deaminases). Cytosine deaminases can catalyze the hydrolytic deamination of cytidine or deoxycytidine to uridine or deoxyuridine, respectively. Thus, in some embodiments, a deaminase or deaminase domain useful with this invention may be a cytidine deaminase domain, catalyzing the hydrolytic deamination of cytosine to uracil. In some embodiments, a cytosine deaminase may be a variant of a naturally-occurring cytosine deaminase, including, but not limited to, a primate (e.g., a human, monkey, chimpanzee, gorilla), a dog, a cow, a rat or a mouse. Thus, in some embodiments, an cytosine deaminase useful with the invention may be about 70% to about 100% identical to a wild-type cytosine deaminase (e.g., about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical, and any range or value therein, to a naturally occurring cytosine deaminase).

[0116] In some embodiments, a cytosine deaminase useful with the invention may be an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the cytosine deaminase may be an APOBEC1 deaminase, an APOBEC2 deaminase, an APOBEC3A deaminase, an APOBEC3B deaminase, an APOBEC3C deaminase, an APOBEC3D deaminase, an APOBEC3F deaminase, an APOBEC3G deaminase, an APOBEC3H deaminase, an APOBEC4 deaminase, a human activation induced deaminase (hAID), an rAPOBEC1, FERNY, and/or a CDA1, optionally a pmCDA1, an atCDA1 (e.g., At2g19570), and evolved versions of the same. Evolved deaminases are disclosed in, for example, U.S. Pat. No. 10,113,163, Gaudelli et al. Nature 551(7681):464-471 (2017)) and Thuronyi et al. (Nature Biotechnology 37: 1070-1079 (2019)), each of which are incorporated by reference herein for their disclosure of deaminases and evolved deaminases. In some embodiments, the cytosine deaminase may be an APOBEC1 deaminase having the amino acid sequence of SEQ ID NO:64. In some embodiments, the cytosine deaminase may be an APOBEC3A deaminase having the amino acid sequence of SEQ ID NO:65. In some embodiments, the cytosine deaminase may be an CDA1 deaminase, optionally a CDA1 having the amino acid sequence of SEQ ID NO:66. In some embodiments, the cytosine deaminase may be a FERNY deaminase, optionally a FERNY having the amino acid sequence of SEQ ID NO:67. In some embodiments, the cytosine deaminase may be a rAPOBEC1 deaminase, optionally a rAPOBEC1 deaminase having the amino acid sequence of SEQ ID NO:68. In some embodiments, the cytosine deaminase may be a hAID deaminase, optionally a hAID having the amino acid sequence of SEQ ID NO:69 or SEQ ID NO:70. In some embodiments, a cytosine deaminase useful with the invention may be about 70% to about 100% identical (e.g., 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or 100% identical) to the amino acid sequence of a naturally occurring cytosine deaminase (e.g., "evolved deaminases") (see, e.g., SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73). In some embodiments, a cytosine deaminase useful with the invention may be about 70% to about 99.5% identical (e.g., about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical) to the amino acid sequence of any one of SEQ ID NOs:64-73 (e.g., at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of any one of SEQ ID NOs:64-73). In some embodiments, a polynucleotide encoding a cytosine deaminase may be codon optimized for expression in a plant and the codon optimized polypeptide may be about 70% to 99.5% identical to the reference polynucleotide.

[0117] An "adenine deaminase" and "adenosine deaminase" as used herein refer to a polypeptide or domain thereof that catalyzes or is capable of catalyzing the hydrolytic deamination (e.g., removal of an amine group from adenine) of adenine or adenosine. In some embodiments, an adenine deaminase may catalyze the hydrolytic deamination of adenosine or deoxyadenosine to inosine or deoxyinosine, respectively. In some embodiments, the adenosine deaminase may catalyze the hydrolytic deamination of adenine or adenosine in DNA. In some embodiments, an adenine deaminase encoded by a nucleic acid construct of the invention may generate an A→G conversion in the sense (e.g., "+"; template) strand of the target nucleic acid or a T→C conversion in the antisense (e.g., "−", complementary) strand of the target nucleic acid. An adenine deaminase useful with this invention may be any known or later identified adenine deaminase from any organism (see, e.g., U.S. Pat. No. 10,113,163, which is incorporated by reference herein for its disclosure of adenine deaminases).

[0118] In some embodiments, an adenosine deaminase may be a variant of a naturally-occurring adenine deaminase. Thus, in some embodiments, an adenosine deaminase may be about 70% to 100% identical to a wild-type adenine deaminase (e.g., about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical, and any range or value therein, to a naturally occurring adenine deaminase). In some embodiments, the deaminase or deaminase does not occur in nature and may be referred to as an engineered, mutated or evolved adenosine deaminase. Thus, for example, an engineered, mutated or evolved adenine deaminase polypeptide or an adenine deaminase domain may be about 70% to 99.9% identical to a naturally occurring adenine deaminase polypeptide/domain (e.g., about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8% or 99.9% identical, and any range or value therein, to a naturally occurring adenine deaminase polypeptide or adenine deaminase domain). In some embodiments, the adenosine deaminase may be from a bacterium, (e.g., *Escherichia coli, Staphylococcus aureus, Haemophilus influenzae, Caulobacter crescentus*, and the like). In some embodiments, a polynucleotide encoding an adenine deaminase polypeptide/domain may be codon optimized for expression in a plant.

[0119] In some embodiments, an adenine deaminase domain may be a wild-type tRNA-specific adenosine deaminase domain, e.g., a tRNA-specific adenosine deaminase (TadA) and/or a mutated/evolved adenosine deaminase domain, e.g., mutated/evolved tRNA-specific adenosine deaminase domain (TadA*). In some embodiments, a TadA domain may be from *E. coli*. In some embodiments, the TadA may be modified, e.g., truncated, missing one or more N-terminal and/or C-terminal amino acids relative to a full-length TadA (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal and/or C terminal amino acid residues may be missing relative to a full length TadA. In some embodiments, a TadA polypeptide or TadA domain does not comprise an N-terminal methionine. In some embodiments, a wild-type *E. coli* TadA comprises the amino acid sequence of SEQ ID NO:74. In some embodiments, a mutated/evolved *E. coli* TadA* comprises the amino acid sequence of any one of SEQ ID NOs:75-78. In some embodiments, a polynucleotide encoding a TadA/TadA* may be codon optimized for expression in a plant. In some embodiments, an adenine deaminase may comprise all or a portion of an amino acid sequence of any one of SEQ ID NOs:79-84. In some embodiments, an adenine deaminase may comprise all or a portion of an amino acid sequence of any one of SEQ ID NOs:74-84.

[0120] In some embodiments, a nucleic acid construct of this invention may further encode a glycosylase inhibitor (e.g., a uracil glycosylase inhibitor (UGI) such as uracil-DNA glycosylase inhibitor). In some embodiments, the invention provides fusion proteins comprising a UGI and/or one or more polynucleotides encoding the same, optionally wherein the one or more polynucleotides may be codon optimized for expression in a plant.

[0121] A "uracil glycosylase inhibitor" useful with the invention may be any protein or polypeptide that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a fragment thereof. In some embodiments, a UGI domain useful with the invention may be about 70% to about 100% identical (e.g., 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or 100% identical and any range or value therein) to the amino acid sequence of a naturally occurring UGI domain. In some embodiments, a UGI domain may comprise the amino acid sequence of SEQ ID NO:85 or a polypeptide having about 70% to about 99.5% identity to the amino acid sequence of SEQ ID NO:85 (e.g., at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of SEQ ID NO:85). For example, in some embodiments, a UGI domain may comprise a fragment of the amino acid sequence of SEQ ID NO:85 that is 100% identical to a portion of consecutive nucleotides (e.g., 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 consecutive nucleotides; e.g., about 10, 15, 20, 25, 30, 35, 40, 45, to about 50, 55, 60, 65, 70, 75, 80 consecutive nucleotides) of the amino acid sequence of SEQ ID NO:85. In some embodiments, a UGI domain may be a variant of a known UGI (e.g., SEQ ID NO:85) having about 70% to about 99.5% identity (e.g., 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% identity, and any range or value therein) to the known UGI. In some embodiments, a polynucleotide encoding a UGI may be codon optimized for expression in a plant (e.g., a plant) and the codon optimized polypeptide may be about 70% to about 99.5% identical to the reference polynucleotide.

[0122] A fusion protein of the present invention may be used in combination with a guide nucleic acid (e.g., guide RNA (gRNA), CRISPR array, CRISPR RNA, crRNA) that is designed to function with the fusion protein to modify a target nucleic acid. A guide nucleic acid useful with this invention may comprise at least one spacer sequence and at least one repeat sequence. The guide nucleic acid is capable of forming a complex with the fusion protein (e.g., with a nuclease domain of the fusion protein) and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the complex to the target nucleic acid, wherein the target nucleic acid may be modified (e.g., cleaved or edited) and/or modulated (e.g., modulating transcription) by a deaminase (e.g., a cytosine deaminase and/or adenine deaminase) or reverse transcriptase, optionally present in and/or recruited to the complex.

[0123] A "guide nucleic acid," "guide RNA," "gRNA," "CRISPR RNA/DNA" "crRNA" or "crDNA" as used herein means a nucleic acid that comprises at least one spacer sequence, which is complementary to (and hybridizes to) a target nucleic acid (e.g., a target DNA and/or protospacer), and at least one repeat sequence (e.g., a repeat of a Type V Cas12a CRISPR-Cas system, or a fragment or portion thereof, wherein the repeat sequence may be linked to the 5' end and/or the 3' end of the spacer sequence. In some embodiments, the guide nucleic acid comprises DNA. In some embodiments, the guide nucleic acid comprises RNA (e.g., is a guide RNA). The design of a gRNA of this invention may be based on a Type I, Type II, Type III, Type IV, Type V, or Type VI CRISPR-Cas system. In some embodiments, a Cas12a gRNA may comprise, from 5' to 3', a repeat sequence (full length or portion thereof ("handle"); e.g., pseudoknot-like structure) and a spacer sequence.

[0124] In some embodiments, a guide nucleic acid may comprise more than one repeat sequence-spacer sequence (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeat-spacer sequences) (e.g., repeat-spacer-repeat, e.g., repeat-spacer-repeat-spacer-repeat-spacer-repeat-spacer-repeat-spacer, and the like). The guide nucleic acids of this invention are synthetic, human-made and not found in nature. A gRNA can be quite long and may be used as an aptamer (like in the MS2 recruitment strategy) or other RNA structures hanging off the spacer.

[0125] A "repeat sequence" as used herein, refers to, for example, any repeat sequence of a wild-type CRISPR Cas locus (e.g., a Cas9 locus, a Cas12a locus, a C2c1 locus, etc.) or a repeat sequence of a synthetic repeat sequence (e.g., a synthetic crRNA) that is functional with the CRISPR-Cas effector protein encoded by the nucleic acid constructs of the invention. A repeat sequence useful with this invention can be any known or later identified repeat sequence of a CRISPR-Cas locus (e.g., Type I, Type II, Type III, Type IV, Type V or Type VI) or it can be a synthetic repeat designed to function in a Type I, II, III, IV, V or VI CRISPR-Cas system. A repeat sequence may comprise a hairpin structure and/or a stem loop structure. In some embodiments, a repeat sequence may form a pseudoknot-like structure at its 5' end (i.e., "handle"). Thus, in some embodiments, a repeat sequence can be identical to or substantially identical to a repeat sequence from wild-type Type I CRISPR-Cas loci, Type II, CRISPR-Cas loci, Type III, CRISPR-Cas loci, Type IV CRISPR-Cas loci, Type V CRISPR-Cas loci and/or Type VI CRISPR-Cas loci. A repeat sequence from a wild-type CRISPR-Cas locus may be determined through established algorithms, such as using the CRISPRfinder offered through CRISPRdb (see, Grissa et al. *Nucleic Acids Res.* 35(Web Server issue):W52-7). In some embodiments, a repeat sequence or portion thereof is linked at its 3' end to the 5' end of a spacer sequence, thereby forming a repeat-spacer sequence (e.g., guide nucleic acid, guide RNA/DNA, crRNA, crDNA).

[0126] In some embodiments, a repeat sequence comprises, consists essentially of, or consists of at least 10 nucleotides depending on the particular repeat and whether the guide nucleic acid comprising the repeat is processed or unprocessed (e.g., 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 to 100 or more nucleotides, or any range or value therein; e.g., about). In some embodiments, a repeat sequence comprises, consists essentially of, or consists of about 10 to about 20, about 10 to about 30, about 10 to about 45, about 10 to about 50, about 15 to about 30, about 15 to about 40, about 15 to about 45, about 15 to about 50, about 20 to about 30, about 20 to about 40, about 20 to about 50, about 30 to about 40, about 40 to about 80, about 50 to about 100 or more nucleotides.

[0127] A repeat sequence linked to the 5' end of a spacer sequence can comprise a portion of a repeat sequence (e.g., 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 or more contiguous nucleotides of a wild-type repeat sequence). In some embodiments, a portion of a repeat sequence linked to the 5' end of a spacer sequence can be about five to about ten consecutive nucleotides in length (e.g., about 5, 6, 7, 8, 9, 10 nucleotides) and have at least 90% sequence identity (e.g., at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more) to the same region (e.g., 5' end) of a wild-type CRISPR Cas repeat nucleotide sequence. In some embodiments, a portion of a repeat sequence may comprise a pseudoknot-like structure at its 5' end (e.g., "handle").

[0128] A "spacer sequence" as used herein is a nucleotide sequence that is complementary to a target nucleic acid (e.g., target DNA) (e.g., protospacer). The spacer sequence can be fully complementary or substantially complementary (e.g., at least about 70% complementary (e.g., about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more)) to a target nucleic acid. Thus, in some embodiments, the spacer sequence can have one, two, three, four, or five mismatches as compared to the target nucleic acid, which mismatches can be contiguous or noncontiguous. In some embodiments, the spacer sequence can have 70% complementarity to a target nucleic acid. In other embodiments, the spacer nucleotide sequence can have 80% complementarity to a target nucleic acid. In still other embodiments, the spacer nucleotide sequence can have 85%, 90%, 95%, 96%, 97%, 98%, 99% or 99.5% complementarity, and the like, to the target nucleic acid (protospacer). In some embodiments, the spacer sequence is 100% complementary to the target nucleic acid. A spacer sequence may have a length from about 15 nucleotides to about 30 nucleotides (e.g., 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides, or any range or value therein). Thus, in some embodiments, a spacer sequence may have complete complementarity or substantial complementarity over a region of a target nucleic acid (e.g., protospacer) that is at least about 15 nucleotides to about 30 nucleotides in length. In some embodiments, the spacer is about 20 nucleotides in length. In some embodiments, the spacer is about 21, 22, or 23 nucleotides in length.

[0129] In some embodiments, the 5' region of a spacer sequence of a guide nucleic acid may be fully complementary to a target nucleic acid, while the 3' region of the spacer may be substantially complementary to the target nucleic acid (such as for a spacer in a Type V CRISPR-Cas system), or the 3' region of a spacer sequence of a guide nucleic acid may be fully complementary to a target nucleic acid, while the 5' region of the spacer may be substantially complementary to the target nucleic acid (such as for a spacer in a Type II CRISPR-Cas system), and therefore, the overall complementarity of the spacer sequence to the target nucleic acid may be less than 100%. Thus, for example, in a guide nucleic acid for a Type V CRISPR-Cas system, the first 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleotides in the 5' region (i.e., seed region) of, for example, a 20 nucleotide spacer sequence may be 100% complementary to the target nucleic acid, while the remaining nucleotides in the 3' region of the spacer sequence are substantially complementary (e.g., at least about 70% complementary) to the target nucleic acid. In some embodiments, the first 1 to 8 nucleotides (e.g., the first 1, 2, 3, 4, 5, 6, 7, 8, nucleotides, and any range therein) of the 5' end of the spacer sequence may be 100% complementary to the target nucleic acid, while the remaining nucleotides in the 3' region of the spacer sequence are substantially complementary (e.g., at least about 50% complementary (e.g., 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more)) to the target nucleic acid.

[0130] As a further example, in a guide nucleic acid for a Type II CRISPR-Cas system, the first 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleotides in the 3' region (i.e., seed region) of, for example, a 20 nucleotide spacer sequence may be 100% complementary to the target nucleic acid, while the remaining nucleotides in the 5' region of the spacer sequence are substantially complementary (e.g., at least about 70% complementary) to the target nucleic acid. In some embodiments, the first 1 to 10 nucleotides (e.g., the first 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleotides, and any range therein) of the 3' end of the spacer sequence may be 100% complementary to the target nucleic acid, while the remaining nucleotides in the 5' region of the spacer sequence are substantially complementary (e.g., at least about 50% complementary (e.g., at least about 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more or any range or value therein)) to the target nucleic acid.

[0131] In some embodiments, a seed region of a spacer may be about 8 to about 10 nucleotides in length, about 5 to about 6 nucleotides in length, or about 6 nucleotides in length.

[0132] In some embodiments, an editing system of the present invention comprises an extended guide nucleic acid, a fusion protein of the present invention, and optionally a reverse transcriptase. In some embodiments, a fusion protein of the present invention comprises all or a portion of a reverse transcriptase. In some embodiments, a fusion protein of the present invention, an extended guide nucleic acid, and optionally a reverse transcriptase may form a complex or may be comprised in a complex that is capable of interacting with a target nucleic acid.

[0133] In some embodiments, a guide nucleic acid further comprises a reverse transcriptase template and may be referred to as an extended guide nucleic acid. An "extended guide nucleic acid" as used herein is a guide nucleic acid as described herein that further comprises a reverse transcriptase template (RTT) and/or a primer binding site (PBS). In some embodiments, an extended guide nucleic acid is an engineered prime editing guide RNA (pegRNA). An extended guide nucleic acid may be a targeted allele guide RNA (tagRNA) or a stabilized targeted allele guide RNA (stagRNA). A "tagRNA" as used herein refers to an extended guide nucleic acid that comprises a PBS and a RTT and has target strand complementarity. A "stagRNA" as used herein refers to a tagRNA that comprises a stabilization motif. A stabilization motif may be present at the 3' and/or 5' end of a tagRNA. In some embodiments, a stabilization motif is present at the 3' end of a tagRNA. Exemplary stabilization motifs include, but are not limited to, recruiting motifs, RNA hairpins, pseudoknot sequences, and/or PP7 motifs (e.g., a PP7 RNA hairpin sequence). In some embodiments, a stagRNA is a tagRNA that comprises a PP7 RNA hairpin sequence. In some embodiments, a CRISPR-Cas effector protein (e.g., a Type II or Type V CRISPR-Cas effector protein), a reverse transcriptase, and an extended guide nucleic acid can form a complex or are comprised in a complex.

[0134] In some embodiments, an extended guide nucleic acid comprises an extended portion that includes a primer binding site and a reverse transcriptase template, wherein the reverse transcriptase template comprises the modification (e.g., edit) to be incorporated into a target nucleic acid. In some embodiments, an extended guide nucleic acid comprises, at its 3' end, a primer binding site and a modification (e.g., an edit) to be incorporated into the target nucleic acid (e.g., a reverse transcriptase template). In some embodiments, an extended guide nucleic acid comprises: (1) a sequence that interacts (e.g., recruits and/or binds) with a CRISPR-Cas effector protein (e.g., a CRISPR-Cas nuclease), (2) a spacer having substantial complementary to a first site on a target nucleic acid (e.g., a CRISPR RNA (crRNA) (a first crRNA) and/or tracrRNA+crRNA (sgRNA)), and (3) a nucleic acid encoded repair template (e.g., an RNA encoded repair template) comprising a primer binding site and an RNA template (e.g., that encodes the modification to be incorporated into the target nucleic acid). In some embodiments, an extended guide nucleic acid (e.g., an extended guide RNA) may comprise, 5'-3', a spacer sequence, a repeat sequence, and an extended portion, the extended portion comprising, 5' to 3', a reverse transcriptase template and a primer binding site. In some embodiments, an extended guide nucleic acid may comprise, 5'-3', a spacer sequence, a repeat sequence and an extended portion, the extended portion comprising, 5' to 3', a primer binding site and a reverse transcriptase template. In some embodiments, an extended guide nucleic acid may comprise, 5'-3', an extended portion, a spacer sequence, and a repeat sequence, wherein the extended portion comprises, 5' to 3', a reverse transcriptase template and a primer binding site. In some embodiments, an extended guide nucleic acid may comprise, 5'-3', an extended portion, a spacer sequence, and a repeat sequence, wherein the extended portion comprises, 5' to 3', a primer binding site and a reverse transcriptase template.

[0135] According to some embodiments, an extended guide nucleic acid (e.g., a pegRNA) may have a structure and/or be designed as described in Anzalone et al., *Nature,* 2019 December; 576(7785): 149-157. In some embodiments, an extended guide nucleic acid comprises a primer binding site (PBS) optionally having a sequence of 1, 2, 3, 4, or 5 to 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides and a reverse transcriptase template (RT template) sequence optionally having a sequence of 65 nucleotides or more. In some embodiments, a PBS of an extended guide nucleic acid has a sequence of less than 15 nucleotides and has a sequence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 nucleotides (e.g., a sequence of 5 or 6 nucleotides in length). The RT template sequence may be after the PBS sequence in the 5' to 3' direction. In some embodiments, the RT template sequence of the extended guide nucleic acid has a length of greater than 65 nucleotides and may comprise about 50 or more nucleotides of heterology relative to the target site (e.g., target nucleic acid), followed by about 15 or more nucleotides of homology relative to the target site. In some embodiments, the RT template sequence of the extended guide nucleic acid is after the PBS sequence and the RT template sequence has a length of greater than 65 nucleotides with the sequence including more than 50 nucleotides of heterology relative to the target site, followed by more than 15 nucleotides of homology relative to the target site. Accordingly, in some embodiments, when the extended guide nucleic acid is reverse transcribed, the resulting newly transcribed sequence may hybridize and/or is configured to hybridize with the unnicked strand of the target site, which may thereby create a heteroduplex DNA with a large insertion into the newly synthesized strand. Upon repair of this mismatched DNA, the resultant repaired DNA may contain a large insertion (e.g., greater than 50 nucleotides) of DNA sequence. In some embodiments, the method may provide a large deletion (e.g., greater than 50 nucleotides) of DNA sequence. In some embodiments, the PBS and the 15 or more nucleotides of homology to the target site may comprise homology arms, which may serve to insert the heterology into the target site optionally using homology directed repair. The inserted DNA may correspond to any functional sequence of DNA such as, but not limited to: a functional transgene; a fragment of DNA that is inserted into a gene in a way that, when the gene is transcribed, would produce a hairpin RNA that is sufficient to silence homologous genes through RNAi; and/or one or more functional site-specific recombination sites, e.g. lox, frt, which could then be used in subsequent Cre or Flp mediated site-specific recombination processes. In some embodiments, an extended guide nucleic acid may be too large to produce using a PolII promoter in vivo. In some embodiments, an extended guide nucleic acid may be operatively associated with and/or produced using a PolII promoter. In some embodiments, a DNA binding polypeptide (e.g., a DNA binding domain) and/or DNA endonuclease may have a structure and/or be designed as described in Anzalone et al., Nature, 2019 December; 576(7785): 149-157. In some embodiments, a DNA binding domain and/or DNA endonuclease is a CRISPR Cas polypeptide such as a Cas9 nickase, a nicking variant of another CRISPR Cas polypeptide, or Cas12a.

[0136] In some embodiments, two extended guide nucleic acids (e.g., pegRNAs) may be used (e.g., an editing system may comprise two extended guide nucleic acids). One or both of the two extended guide nucleic acids may have a structure and/or be designed as described in Anzalone et al., Nature, 2019 December; 576(7785): 149-157. The two extended guide nucleic acids may comprise a primer binding site (PBS) optionally having a sequence of 1, 2, 3, 4, or 5 to 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides and a reverse transcriptase template (RT template) sequence optionally having a sequence of 50 nucleotides or more. The RT template sequences of the two extended guide nucleic acids may be complementary to each other and as such the polynucleotides that are respectively reverse transcribed from each the RT templates will be complementary to each other and will be able to hybridize with each other. This may allow for the intermediates that are produced by this system and/or method to join together two sections of DNA that are otherwise separated by more than 50 nucleotides, e.g. within a chromosome, or that are positioned on two separate pieces of DNA, e.g. on two different chromosomes. After repair of the intermediates, the resultant products may produce, depending on the design of the RT template, large deletions, large inversions, or inter-chromosomal recombinations. Since all of these products are produced by homology directed repair, the products may be predictably precise and/or reproducible. In some embodiments, a DNA binding polypeptide (e.g., a DNA binding domain) and/or DNA endonuclease may have a structure and/or be designed as described in Anzalone et al., Nature, 2019 December; 576 (7785): 149-157. In some embodiment, a DNA binding polypeptide and/or DNA endonuclease is a CRISPR Cas polypeptide such as a Cas9 nickase, a similar nicking variant of another CRISPR Cas polypeptide, or Cas12a. In some embodiments, a DNA binding polypeptide and/or DNA endonuclease is a Cas9 nuclease, a similar nuclease from another CRISPR Cas polypeptide, or Cas12a. Using a nuclease (rather than a nickase) may facilitate the intra- or interchromosomal recombination processes through single-strand annealing of the more than 50 nucleotide 3' overhangs that would be produced at each of the two target sites corresponding to the two pegRNA target nucleic acids. In some embodiments, an editing system comprises one extended guide nucleic acid and a guide nucleic acid that is devoid of a reverse transcriptase template and/or primer binding site.

[0137] An extended guide nucleic acid may comprise a CRISPR nucleic acid (e.g., CRISPR RNA, CRISPR DNA, crRNA, crDNA) and/or a CRISPR nucleic acid and a tracr nucleic acid; and (b) an extended portion comprising a primer binding site and a reverse transcriptase template (RT template), wherein the RT template encodes a modification to be incorporated into the target nucleic acid. The CRISPR nucleic acid may be a Type II or Type V CRISPR nucleic acid and/or the tracr nucleic acid may be any tracr corresponding to the appropriate Type II or Type V CRISPR nucleic acid. In some embodiments, an extended guide nucleic acid comprises: (i) a Type V CRISPR nucleic acid or a Type II CRISPR nucleic acid (e.g., a Type II or Type V CRISPR RNA, Type II or Type V CRISPR DNA, Type II or Type V crRNA, or Type II or Type V crDNA) and/or a CRISPR nucleic acid and a tracr nucleic acid (e.g., a Type II or Type V tracrRNA, Type II or Type V tracrDNA); and (ii) an extended portion comprising a primer binding site and a reverse transcriptase template (RT template), wherein the Type V CRISPR nucleic acid or Type II CRISPR nucleic acid comprises a spacer that binds to a first strand (e.g., the target strand) of a target nucleic acid (e.g., the spacer is complementary to a portion of consecutive nucleotides in the first strand of the target nucleic acid) and the primer binding site binds to the first strand (e.g., target strand). In some embodiments, the extended portion can be fused to either the 5' end or 3' end of the CRISPR nucleic acid (e.g., from 5' to 3': repeat-spacer-extended portion or extended portion-repeat-spacer) and/or to the 5' or 3' end of the tracr nucleic acid. In some embodiments, the extended portion of an extended guide nucleic acid comprises, 5' to 3', an RT template (RTT) and a primer binding site (PBS) (e.g., 5'-crRNA-spacer-RTT(edit encoded)-PBS-3') or comprises 5' to 3' a PBS and RTT, depending on the location of the extended portion relative to the CRISPR nucleic acid of the extended guide nucleic acid (e.g., 5'-crRNA-spacer-PBS-RTT(edit encoded)-3'). For example, in some embodiments, an extended portion of the extended guide nucleic acid may comprise, 5' to 3', an RT template and a primer binding site (when the extended guide is linked to the 3' end of the CRISPR nucleic acid). In some embodiments, an extended portion of the extended guide may comprise, 5' to 3', a primer binding site and an RT template (when the extended guide is linked to the 5' end of the CRISPR nucleic acid).

[0138] In some embodiments, a target nucleic acid is double stranded and comprises a first strand and a second strand and a primer binding site of an extended guide nucleic acid binds to the second strand (e.g., the non-target, top strand) of the target nucleic acid. In some embodiments, a target nucleic acid is double stranded and comprises a first strand and a second strand and a primer binding site of an extended guide nucleic acid binds to the first strand (e.g., binds to the target strand, optionally the same strand to which a CRISPR-Cas effector protein is recruited, bottom strand) of the target nucleic acid. In some embodiments, a target nucleic acid is double stranded and comprises a first strand and a second strand and the primer binding site of an extended guide nucleic acid binds to the second strand (e.g., the non-target strand, optionally the opposite strand from that to which the CRISPR-Cas effector protein is recruited) of the target nucleic acid. In some embodiments, a reverse transcriptase (RT) may add to the target strand of a target nucleic acid (e.g., the strand to which the spacer of the CRISPR nucleic acid of the extended guide nucleic acid is complementary and to which the CRISPR-Cas effector protein is recruited). In some embodiments, the reverse transcriptase (RT) adds to the non-target strand of a target nucleic acid (e.g., the strand that is complementary to the strand to which the spacer of the CRISPR nucleic acid is complementary and to which the CRISPR-Cas effector protein is recruited). Example methods and editing systems are described in International Patent Publication No. WO 2021/092130, International Patent Publication No. WO 2022/098993, and U.S. Patent Application Publication Nos. 2021/0147862, 2021/0130835, 2021/0147862, and 2022/0145334, each of which are incorporated herein by reference in their entirety.

[0139] The RT template of an extended guide nucleic acid may encode one or more modification(s) (e.g., edit(s)) to be incorporated into a target nucleic acid. The one or more modification(s) may be located in any position within an RT template (e.g., where the position location may be relative to the position of a protospacer adjacent motif (PAM) of the target nucleic acid). In some embodiments, an RT template has a modification at one or more positions from −1 to 23 (e.g., −1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, or 23) relative to the position of a protospacer adjacent motif (PAM) (e.g., TTTG) in a target nucleic acid. In some embodiments, an RT template may comprise a modification located at nucleotide position −1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, or 23. In some embodiments, an RT template may comprise a modification located at nucleotide position 4 to nucleotide position 17 (e.g., position 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17) of the RT template relative to the position of a PAM of a target nucleic acid. In some embodiments, an RT template may comprise a modification located at nucleotide position 10 to nucleotide position 17 (e.g., position 10, 11, 12, 13, 14, 15, 16, or 17) of the RT template relative to the position of a PAM of a target nucleic acid. In some embodiments, an RT template may comprise a modification located at nucleotide position 12 to nucleotide position 15 (e.g., position 12, 13, 14, or 15) of the RT template relative to the position of a PAM of a target nucleic acid.

[0140] In some embodiments, an extended portion of an extended guide nucleic acid may comprise, 5' to 3', an RT template and a primer binding site (e.g., when the extended portion is linked to the 3' end of a CRISPR nucleic acid). In some embodiments, an extended portion of an extended guide nucleic acid may comprise, 5' to 3', a primer binding site and an RT template (RTT) (e.g., when the extended portion is linked to the 5' end of the CRISPR nucleic acid). In some embodiments, an RT template may have a length of about 1 nucleotide to about 100 nucleotides (e.g., about 1, 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, 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleotides, and any range or value therein), e.g., about 1 nucleotide to about 10 nucleotides, about 1 nucleotide to about 15 nucleotides, about 1 nucleotide to about 20 nucleotides, about 1 nucleotide to about 25 nucleotides, about 1 nucleotide to about 30 nucleotides, about 1 nucleotide to about 35, 36, 37, 38, 39 or 40 nucleotides, about 1 nucleotide to about 50 nucleotides, about 5 nucleotides to about 15 nucleotides, about 5 nucleotides to about 20 nucleotides, about 5 nucleotides to about 25 nucleotides, about 5 nucleotides to about 30 nucleotides, about 5 nucleotides to about 35, 36, 37, 38, 39 or 40 nucleotides, about 5 nucleotides to about 50 nucleotides, about 8 nucleotides to about 15 nucleotides, about 8 nucleotide to about 20 nucleotides, about 8 nucleotide to about 25 nucleotides, about 8 nucleotide to about 30 nucleotides, about 8 nucleotide to about 35, 36, 37, 38, 39 or 40 nucleotides, about 8 nucleotide to about 50 nucleotides in length, about 8 nucleotides to about 100 nucleotides, about 10 nucleotide to about 15 nucleotides, about 10 nucleotide to about 20 nucleotides, about 10 nucleotide to about 25 nucleotides, about 10 nucleotide to about 30 nucleotides, about 10 nucleotide to about 36 nucleotides, about 10 nucleotide to about 40 nucleotides, about 10 nucleotide to about 50 nucleotides, about 10 nucleotides to about 100 nucleotides in length and any range or value therein. In some embodiments, the length of an RT template may be at least 8 nucleotides, optionally about 8 nucleotides to about 100 nucleotides. In some embodiments, the length of an RT template is 36, 37, 38, 39 or 40 nucleotides or less (e.g., about 1, 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, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 nucleotides in length, or any value or range therein (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 nucleotides in length to about 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 nucleotides in length). In some embodiments, the length of an RT template may be at least 30 nucleotides, optionally about 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 nucleotides in length to about to about 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or 80 nucleotides in length, or any range or value therein. In some embodiments, the length of an RT template may be about 36, 40, 44, 47, 50, 52, 55, 63, 72 or 74 nucleotides.

[0141] Within the length of the RTT one or more modification(s) may be present. The one or more modification(s) may be located anywhere within the RTT, wherein the position of the modification may be described relative to the position of a protospacer adjacent motif (PAM) of a target nucleic acid. In some embodiments, an RT template may comprise a modification located at nucleotide position −1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, or 23. In some embodiments, an RT template may comprise a modification located at nucleotide position 4 to nucleotide position 17 (e.g., position 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17) of the RT template relative to the position of a protospacer adjacent motif (PAM) of a target nucleic acid. In some embodiments, an RT template may comprise a modification located at nucleotide position 10 to nucleotide position 17 (e.g., position 10, 11, 12, 13, 14, 15, 16, or 17) of the RT template relative to the position of a protospacer adjacent motif (PAM) of a target nucleic acid. In some embodiments, an RT template may comprise a modification located at nucleotide position 12 to nucleotide position 15 (e.g., position 12, 13, 14, or 15) of the RT template relative to the position of a protospacer adjacent motif (PAM) of a target nucleic acid.

[0142] As used herein, a “primer binding site” (PBS) of an extended portion of an extended guide nucleic acid (e.g., a tagRNA) refers to a sequence of consecutive nucleotides that can bind to a region or “primer” on a target nucleic acid, e.g., is complementary to the target nucleic acid primer. As an example, a CRISPR Cas effector protein (e.g., a Type II or Type V, e.g., Cas 9 or Cas12a) may nick/cut the DNA and the 3' end of the cut DNA acts as a primer for the PBS portion of the extended guide nucleic acid. The PBS may be complementary to the 3' end of a strand of the target nucleic acid and may bind and/or may be configured to bind to either the target strand or non-target strand. A primer binding site can be fully complementary to the primer or it may be substantially complementary (e.g., at least 70% complementary (e.g., 70% or about 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more)) to the primer of a target nucleic acid. In some embodiments, the length of a primer binding site of an extended portion may be about 1 nucleotide to about 100 nucleotides in length (e.g., about 1, 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, 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleotides, or any value or range therein), or about 4 nucleotide to about 85 nucleotides, about 10 nucleotide to about 80 nucleotides, about 20 nucleotide to about 80 nucleotides, about 25 nucleotides to about 80 nucleotides about 30 nucleotide to about 80 nucleotides, about 40 nucleotide to about 80 nucleotides, about 45 nucleotide to about 80 nucleotides, about 45 nucleotide to about 75 nucleotides, or about 45 nucleotide to about 60 nucleotides, or any range or value therein. In some embodiments, the length of a PBS may be at least 30 nucleotides, optionally about 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 nucleotides to about 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or 80 nucleotides in length, or any range or value therein. In some embodiments, the length of a PBS may be about 8, 16, 24, 32, 40, 48, 56, 64, 72, or 80 nucleotides.

[0143] In some embodiments, an RTT may have a length of about 35 nucleotides to about 75 nucleotides and a PBS may have a length of about 30 nucleotides to about 80 nucleotides, optionally wherein the PBS may comprise a length of about 8, 16, 24, 32, 40, 48, 56, 64, 72, or 80 nucleotides and the RTT may comprise a length of about 36, 40, 44, 47, 50, 52, 55, 63, 72 or 74 nucleotides, or any combination thereof of the RTT length and/or PBS length.

[0144] In some embodiments, an extended portion of an extended guide nucleic acid may be fused to either the 5' end or 3' end of a Type II or a Type V CRISPR nucleic acid (e.g., 5' to 3': repeat-spacer-extended portion, or extended portion-repeat-spacer) and/or to the 5' or 3' end of the tracr nucleic acid. In some embodiments, when an extended portion is located 5' of the crRNA, a Type V CRISPR-Cas effector protein is modified to reduce (or eliminate) self-processing RNAse activity.

[0145] In some embodiments, the extended portion of an extended guide nucleic acid may be linked to the Type II or Type V CRISPR nucleic acid and/or the Type II or Type V tracrRNA via a linker. In some embodiments, a linker have a length of about 1 to about 100 nucleotides or more (e.g., about 1, 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, 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleotides in length, and any range therein (e.g., about 2 to about 40, about 2 to about 50, about 2 to about 60, about 4 to about 40, about 4 to about 50, about 4 to about 60, about 5 to about 40, about 5 to about 50, about 5 to about 60, about 9 to about 40, about 9 to about 50, about 9 to about 60, about 10 to about 40, about 10 to about 50, about 10 to about 60, about 40 to about 100, about 50 to about 100, or about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 nucleotides to about 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, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleotides in length (e.g., about 105, 110, 115, 120, 130, 140 150 or more nucleotides in length).

[0146] A guide nucleic acid and/or an extended guide nucleic acid may comprise one or more recruiting motifs as described herein, which may be linked to the 5' end and/or the 3' end of the guide nucleic acid and/or it may be inserted into the guide nucleic acid (e.g., within a hairpin loop of the guide nucleic acid). In some embodiments, an extended guide nucleic acid may be linked to an RNA recruiting motif. An extended guide nucleic acid and/or guide nucleic acid may be linked to one or to two or more RNA recruiting motifs (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more motifs; e.g., at least 10 to about 25 motifs), optionally wherein the two or more RNA recruiting motifs may be the same RNA recruiting motif or different RNA recruiting motifs. In some embodiments, an RNA recruiting motif may be located on the 3' end of the extended portion of an extended guide nucleic acid (e.g., 5'-3', repeat-spacer-extended portion (RT template-primer binding site)-RNA recruiting motif). In some embodiments, an RNA recruiting motif may be embedded in the extended portion of an extended guide nucleic acid.

[0147] In some embodiments, an editing system comprises an extended guide nucleic acid that is linked to an RNA recruiting motif and a reverse transcriptase that is a reverse transcriptase fusion protein, wherein the reverse transcriptase fusion protein comprises a reverse transcriptase polypeptide fused to an affinity polypeptide that binds to the RNA recruiting motif, wherein the extended guide nucleic acid binds to a target nucleic acid and the RNA recruiting motif binds to the affinity polypeptide, thereby recruiting the reverse transcriptase fusion protein to the extended guide nucleic acid and contacting the target nucleic acid with the reverse transcriptase. In some embodiments, two or more reverse transcriptase fusion proteins may be recruited to an extended guide nucleic acid, thereby contacting the target nucleic acid with two or more reverse transcriptase fusion proteins.

[0148] A "target nucleic acid", "target DNA," "target nucleotide sequence," "target region," and "target region in the genome" are used interchangeably herein and refer to a region of an organism's (e.g., a plant's) genome that comprises a sequence that is fully complementary (100% complementary) or substantially complementary (e.g., at least 70% complementary (e.g., 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more)) to a spacer sequence in a guide nucleic acid as defined herein. A target nucleic acid is targeted by an editing system (or a component thereof) as described herein. A target region useful for a CRISPR-Cas system may be located immediately 3' (e.g., Type V CRISPR-Cas system) or immediately 5' (e.g., Type II CRISPR-Cas system) to a PAM sequence in the genome of the organism (e.g., a plant genome or mammalian (e.g., human) genome). A target region may be selected from any region of at least 15 consecutive nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 nucleotides, and the like) located immediately adjacent to a PAM sequence.

[0149] A "protospacer sequence" or "protospacer" as used herein refer to a sequence that is fully or substantially complementary to (and can hybridize to) a spacer sequence of a guide nucleic acid. In some embodiments, the protospacer is all or a portion of a target nucleic acid as defined herein that is fully or substantially complementary (and hybridizes) to the spacer sequence of the CRISPR repeat-spacer sequences (e.g., guide nucleic acids, CRISPR arrays, crRNAs).

[0150] In the case of Type V CRISPR-Cas (e.g., Cas12a) systems and Type II CRISPR-Cas (Cas9) systems, the protospacer sequence is flanked by (e.g., immediately adjacent to) a protospacer adjacent motif (PAM). For Type IV CRISPR-Cas systems, the PAM is located at the 5' end on the non-target strand and at the 3' end of the target strand (see below, as an example).

```
5'-NNNNNNNNNNNNNNNNNNNN-3' RNA Spacer
  ||||||||||||||||||||
3'AAANNNNNNNNNNNNNNNNNNNN-5' Target strand
  ||||
5'TTTNNNNNNNNNNNNNNNNNNNN-3' Non-target strand
```

[0151] In the case of Type II CRISPR-Cas (e.g., Cas9) systems, the PAM is located immediately 3' of the target region. The PAM for Type I CRISPR-Cas systems is located 5' of the target strand. There is no known PAM for Type III CRISPR-Cas systems. Makarova et al. describes the nomenclature for all the classes, types and subtypes of CRISPR systems (Nature Reviews Microbiology 13:722-736 (2015)). Guide structures and PAMs are described in R. Barrangou (*Genome Biol.* 16:247 (2015)).

[0152] Canonical Cas12a PAMs are T rich. In some embodiments, a canonical Cas12a PAM sequence may be 5'-TTN, 5'-TTTN, or 5'-TTTV. In some embodiments, canonical Cas9 (e.g., *S. pyogenes*) PAMs may be 5'-NGG-3'. In some embodiments, non-canonical PAMs may be used but may be less efficient.

[0153] Additional PAM sequences may be determined by those skilled in the art through established experimental and computational approaches. Thus, for example, experimental approaches include targeting a sequence flanked by all possible nucleotide sequences and identifying sequence members that do not undergo targeting, such as through the transformation of target plasmid DNA (Esvelt et al. 2013. *Nat. Methods* 10:1116-1121; Jiang et al. 2013. *Nat. Biotechnol.* 31:233-239). In some aspects, a computational approach can include performing BLAST searches of natural spacers to identify the original target DNA sequences in bacteriophages or plasmids and aligning these sequences to determine conserved sequences adjacent to the target sequence (Briner and Barrangou. 2014. *Appl. Environ. Microbiol.* 80:994-1001; Mojica et al. 2009. *Microbiology* 155:733-740).

[0154] In some embodiments, the present invention provides expression cassettes and/or vectors comprising the nucleic acid constructs of the invention (e.g., one or more components of an editing system of the invention). In some embodiments, expression cassettes and/or vectors comprising the nucleic acid constructs of the invention and/or one or more guide nucleic acids may be provided. In some embodiments, a nucleic acid construct of the invention encodes a fusion protein, and/or a deaminase, and each may be comprised on the same or on a separate expression cassette or vector from that comprising the one or more guide nucleic acids. When the nucleic acid construct encoding a fusion protein or the components of an editing system is/are comprised on separate expression cassette(s) or vector(s) from that comprising the guide nucleic acid, a target nucleic acid may be contacted with (e.g., provided with) the expression cassette(s) or vector(s) encoding the fusion protein or components of an editing system in any order from one another and the guide nucleic acid, e.g., prior to, concurrently with, or after the expression cassette comprising the guide nucleic acid is provided (e.g., contacted with the target nucleic acid).

[0155] Methods of recruiting one or more components of an editing system to each other and/or to a target nucleic acid are known in the art and may include the use of a peptide tag or an affinity polypeptide that interacts with the peptide tag. In some embodiments, a guide nucleic acid may be linked to an RNA recruiting motif and a deaminase may be linked to an affinity polypeptide capable of interacting with the RNA recruiting motif, thereby recruiting the deaminase to the target nucleic acid. Alternatively, chemical interactions may be used to recruit a polypeptide (e.g., a deaminase) to a target nucleic acid.

[0156] A "recruiting motif" as used herein refers to one half of a binding pair that may be used to recruit a compound to which the recruiting motif is bound to another compound that includes the other half of the binding pair (i.e., a "corresponding motif"). The recruiting motif and corresponding motif may bind noncovalently. In some embodiments, a recruiting motif is an RNA recruiting motif (e.g., an RNA recruiting motif that is capable of binding and/or configured to bind to an affinity polypeptide), an affinity polypeptide (e.g., an affinity polypeptide that is capable of binding and/or configured to bind an RNA recruiting motif and/or a peptide tag), or a peptide tag (e.g., a peptide tag that is capable of binding and/or configured to bind an affinity polypeptide). For example, when a recruiting motif is an RNA recruiting motif, the corresponding motif for the RNA recruiting motif may be an affinity polypeptide that binds the RNA recruiting motif. A further example is that when a recruiting motif is a peptide tag, the corresponding motif for the peptide tag may be an affinity polypeptide that binds the peptide tag. Thus, a compound comprising a recruiting motif (e.g., an affinity polypeptide) may be recruited to another compound (e.g., a guide nucleic acid) comprising a corresponding motif for the recruiting motif (e.g., an RNA recruiting motif).

[0157] A peptide tag (e.g., epitope) useful with this invention may include, but is not limited to, a GCN4 peptide tag (e.g., Sun-Tag), a c-Myc affinity tag, an HA affinity tag, a His affinity tag, an S affinity tag, a methionine-His affinity tag, an RGD-His affinity tag, a FLAG octapeptide, a strep tag or strep tag II, a V5 tag, and/or a VSV-G epitope. Any epitope that may be linked to a polypeptide and for which there is a corresponding affinity polypeptide that may be linked to another polypeptide may be used with this invention as a peptide tag. In some embodiments, a peptide tag may comprise 1 or 2 or more copies of a peptide tag (e.g., repeat unit, multimerized epitope (e.g., tandem repeats)) (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more repeat units. In some embodiments, an affinity polypeptide that interacts with/binds to a peptide tag may be an antibody. In some embodiments, the antibody may be a scFv antibody. In some embodiments, an affinity polypeptide that binds to a peptide tag may be synthetic (e.g., evolved for affinity interaction) including, but not limited to, an affibody, an anticalin, a monobody and/or a DARPin (see, e.g., Sha et al., *Protein Sci.* 26(5): 910-924 (2017)); Gilbreth (*Curr Opin Struc Biol* 22(4):413-420 (2013)), U.S. Pat. No. 9,982,053, each of which are incorporated by reference in their entireties for the teachings relevant to affibodies, anticalins, monobodies and/or DARPins.

[0158] In some embodiments, a guide nucleic acid may be linked to an RNA recruiting motif, and a polypeptide to be recruited (e.g., a deaminase) may be fused to an affinity polypeptide that binds to the RNA recruiting motif, wherein the guide binds to the target nucleic acid and the RNA recruiting motif binds to the affinity polypeptide, thereby recruiting the polypeptide to the guide and contacting the target nucleic acid with the polypeptide (e.g., deaminase). In some embodiments, two or more polypeptides may be recruited to a guide nucleic acid, thereby contacting the target nucleic acid with two or more polypeptides (e.g., deaminases).

[0159] In some embodiments of the invention, a guide RNA may be linked to one or to two or more RNA recruiting motifs (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more motifs; e.g., at least 10 to about 25 motifs), optionally wherein the two or more RNA recruiting motifs may be the same RNA recruiting motif or different RNA recruiting motifs. In some embodiments, an RNA recruiting motif and corresponding affinity polypeptide may include, but is not limited, to a telomerase Ku binding motif (e.g., Ku binding hairpin) and the corresponding affinity polypeptide Ku (e.g., Ku heterodimer), a telomerase Sm7 binding motif and the corresponding affinity polypeptide Sm7, an MS2 phage operator stem-loop and the corresponding affinity polypeptide MS2 Coat Protein (MCP), a PP7 phage operator stem-loop and the corresponding affinity polypeptide PP7 Coat Protein (PCP), an SfMu phage Com stem-loop and the corresponding affinity polypeptide Com RNA binding protein, a PUF binding site (PBS) and the affinity polypeptide Pumilio/fem-3 mRNA binding factor (PUF), and/or a synthetic RNA-aptamer and the aptamer ligand as the corresponding affinity polypeptide. In some embodiments, the RNA recruiting motif and corresponding affinity polypeptide may be an MS2 phage operator stem-loop and the affinity polypeptide MS2 Coat Protein (MCP). In some embodiments, the RNA recruiting motif and corresponding affinity polypeptide may be a PUF binding site (PBS) and the affinity polypeptide Pumilio/fem-3 mRNA binding factor (PUF). Exemplary RNA recruiting motifs and corresponding affinity polypeptides that may be useful with this invention can include, but are not limited to, SEQ ID NOs:86-96.

[0160] In some embodiments, the components for recruiting polypeptides and nucleic acids may include those that function through chemical interactions that may include, but are not limited to, rapamycin-inducible dimerization of FRB-FKBP; Biotin-streptavidin; SNAP tag; Halo tag; CLIP tag; DmrA-DmrC heterodimer induced by a compound; bifunctional ligand (e.g., chemically induced dimerization).

[0161] As described herein, a "peptide tag" may be employed to recruit one or more polypeptides. A peptide tag may be any polypeptide that is capable of being bound by a corresponding motif such as an affinity polypeptide. A peptide tag may also be referred to as an "epitope" and when provided in multiple copies, a "multimerized epitope." Example peptide tags can include, but are not limited to, a GCN4 peptide tag (e.g., Sun-Tag), a c-Myc affinity tag, an HA affinity tag, a His affinity tag, an S affinity tag, a methionine-His affinity tag, an RGD-His affinity tag, a FLAG octapeptide, a strep tag or strep tag II, a V5 tag, and/or a VSV-G epitope. In some embodiments, a peptide tag may also include phosphorylated tyrosines in specific sequence contexts recognized by SH2 domains, characteristic consensus sequences containing phosphoserines recognized by 14-3-3 proteins, proline rich peptide motifs recognized by SH3 domains, PDZ protein interaction domains or the PDZ signal sequences, and an AGO hook motif from plants. Peptide tags are disclosed in WO2018/136783 and U.S. Patent Application Publication No. 2017/0219596, which are incorporated by reference for their disclosures of peptide tags. Peptide tags that may be useful with this invention can include, but are not limited to, SEQ ID NO:97 and SEQ ID NO:98. An affinity polypeptide useful with peptide tags includes, but is not limited to, SEQ ID NO:99.

[0162] A peptide tag may comprise or be present in one copy or in 2 or more copies of the peptide tag (e.g., multimerized peptide tag or multimerized epitope) (e.g., about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 9, 20, 21, 22, 23, 24, or 25 or more peptide tags). When multimerized, the peptide tags may be fused directly to one another or they may be linked to one another via one or more amino acids (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more amino acids, optionally about 3 to about 10, about 4 to about 10, about 5 to about 10, about 5 to about 15, or about 5 to about 20 amino acids, and the like, and any value or range therein. Thus, in some embodiments, a CRISPR-Cas effector protein of the invention may comprise a CRISPR-Cas effector protein fused to one peptide tag or to two or more peptide tags, optionally wherein the two or more peptide tags are fused to one another via one or more amino acid residues. In some embodiments, a peptide tag useful with the invention may be a single copy of a GCN4 peptide tag or epitope or may be a multimerized GCN4 epitope comprising about 2 to about 25 or more copies of the peptide tag (e.g., about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more copies of a GCN4 epitope or any range therein).

[0163] In some embodiments, a peptide tag may be fused to a CRISPR-Cas polypeptide or domain. In some embodiments, a peptide tag may be fused or linked to the C-terminus of a CRISPR-Cas effector protein to form a CRISPR-Cas fusion protein. In some embodiments, a peptide tag may be fused or linked to the N-terminus of a CRISPR-Cas effector protein to form a CRISPR-Cas fusion protein. In some embodiments, a peptide tag may be fused within a CRISPR-Cas effector protein (e.g., a peptide tag may be in a loop region of a CRISPR-Cas effector protein). In some embodiments, peptide tag may be fused to a cytosine deaminase and/or to an adenine deaminase.

[0164] An "affinity polypeptide" (e.g., "recruiting polypeptide") refers to any polypeptide that is capable of binding to its corresponding peptide tag, peptide tag, or RNA recruiting motif. An affinity polypeptide for a peptide tag may be, for example, an antibody and/or a single chain antibody that specifically binds the peptide tag, respectively. In some embodiments, an antibody for a peptide tag may be, but is not limited to, an scFv antibody. In some embodiments, an affinity polypeptide may be fused or linked to the N-terminus of a deaminase (e.g., a cytosine deaminase or an adenine deaminase). In some embodiments, the affinity polypeptide is stable under the reducing conditions of a cell or cellular extract.

[0165] The nucleic acid constructs of the invention and/or guide nucleic acids may be comprised in one or more expression cassettes as described herein. In some embodiments, a nucleic acid construct of the invention may be comprised in the same or in a separate expression cassette or vector from that comprising a guide nucleic acid and/or an extended guide nucleic acid.

[0166] In some embodiments, a nucleic acid construct, expression cassette, or vector of the invention that is optimized for expression in an organism (e.g., a human or plant) may be about 70% to 100% identical (e.g., about 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or 100%) to a nucleic acid construct, expression cassette or vector comprising the same polynucleotide(s) but which have not been codon optimized for expression in the organism.

[0167] When used in combination a guide nucleic acid, a nucleic acid construct of the invention (and expression cassette and/or vector comprising the same) may be used to modify a target nucleic acid and/or its expression. A target nucleic acid may be contacted with a nucleic acid construct of the invention and/or expression cassettes and/or vectors comprising the same prior to, concurrently with or after contacting the target nucleic acid with the guide nucleic acid/recruiting guide nucleic acid (and/or expression cassettes and vectors comprising the same.

[0168] According to embodiments of the present invention, provided herein are fusion proteins (e.g., engineered proteins) that include an intein polypeptide. In some embodiments, a fusion protein of the present invention includes a Cas12a polypeptide and an intein polypeptide. In some embodiments, a fusion protein of the present invention includes a polypeptide of interest and an intein polypeptide. In some embodiments, a fusion protein of the present invention includes a reverse transcriptase polypeptide and an intein polypeptide. An "engineered protein" as used herein is a polypeptide or protein that is not found naturally in nature. A fusion protein of the present invention may comprise a Cas12a polypeptide and/or a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to (e.g., linked and/or attached to) an intein polypeptide. A Cas12a polypeptide and an intein polypeptide may be directly fused (e.g., no amino acid residue or linker between the two polypeptides) or indirectly fused (e.g., a linker (e.g., an amino acid or peptide) or another polypeptide is between the two polypeptides). Similarly, a polypeptide of interest (e.g., a reverse transcriptase polypeptide) and an intein polypeptide may be directly fused or indirectly fused. In some embodiments, a Cas12a polypeptide is directly fused (e.g., via a peptide bond) to an intein polypeptide. In some embodiments, a Cas12a polypeptide is indirectly fused (e.g., via a peptide linker) to an intein polypeptide. A Cas12a polypeptide and an intein polypeptide may be fused in any orientation. For example, in some embodiments, the N-terminus of the Cas12a polypeptide is fused to the C-terminus of the intein polypeptide or to the N-terminus of the intein polypeptide. In some embodiments, the C-terminus of the Cas12a polypeptide is fused to the C-terminus of the intein polypeptide or to the N-terminus of the intein polypeptide. A polypeptide of interest (e.g., a reverse transcriptase polypeptide) and an intein polypeptide may be fused in any orientation. In some embodiments, a fusion protein of the present invention has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to one or more of SEQ ID NOs:100-109 or 187-188. According to some embodiments, provided is a nucleic acid molecule that encodes a polypeptide that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more of SEQ ID NOs:100-109 or 187-188. In some embodiments, a nucleic acid molecule comprises a polynucleotide that has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more of SEQ ID NOs:185-186.

[0169] In some embodiments, an intein polypeptide is a portion (e.g., a fragment such as a N-terminal intein fragment or a C-terminal intein fragment) of an intein such as a portion of a molecular scaffold formed from two corresponding portions (e.g., two corresponding fragments or a pair of intein polypeptides) that together can or are configured to catalyze both the cleavage and formation of a peptide bond. An "intein" as used herein refers to a catalytically active complex formed from two polypeptides (e.g., a pair of intein polypeptides) that are associated with each other, wherein the complex can or is configured to excise itself (e.g., the two polypeptides) from a larger precursor polypeptide and can or is configured to ligate the ends flanking the two polypeptides with a peptide bond, optionally wherein the excising and ligating occur concurrently. In some embodiments, an intein polypeptide is a portion of a split intein such as a trans-splicing splint intein. A "split intein" as used herein can perform protein trans-splicing in which two fragments of the intein (e.g., two intein polypeptides or a pair of intein polypeptides) associate (e.g., non-covalently bind) to form a catalytically competent complex or molecular scaffold that catalyzes excision of the two intein fragments and the ligation of their flanking sequences. In some embodiments, an intein polypeptide is an autocatalytic polypeptide that together with a corresponding intein polypeptide to form an intein (e.g., a split intein) is capable of excising the intein polypeptide from a larger precursor protein (e.g., a fusion protein of the present invention) and enable the flanking polypeptide sequences (e.g., the sequence adjacent to the excised intein polypeptide) to be ligated through the formation of a new peptide bond. A fusion protein of the present invention may include an intein polypeptide that is one part of two total parts such that the intein polypeptide together with another intein polypeptide (e.g., the second part) together form an intein such as a trans-splicing split intein. In some embodiments, a split intein and/or an intein polypeptide thereof may be able to function (e.g., perform protein trans-splicing) without any assistance and/or conditions other than the two portions of the split intein (e.g., the two intein polypeptides that together provide the split intein). For example, two intein polypeptides may spontaneously associate to form the intein and may spontaneously catalyze their own excision and the ligation of their flanking sequences without assistance (e.g., an external condition and/or cofactor). In some embodiments, an intein may be used for which protein trans-splicing is controlled (e.g., the intein undergoes conditional trans-splicing). For example, certain conditions (e.g., light and/or a cofactor) may be required for an intein to function. In some embodiments, a split intein and/or an intein polypeptide is a light inducible intein (e.g., as described in Wong S, et al. (2015) An Engineered Split Intein for Photoactivated Protein Trans-Splicing. PLoS ONE 10(8): e0135965) that uses light in order to control the association of the two intein polypeptides that together provide the intein (e.g., the catalytically active complex). In some embodiments, a cofactor (e.g., a small molecule) and/or activator is used to bring together two intein polypeptides that together provide the intein and thereby control protein trans-splicing such as described in Gramespacher, Josef A., et al. *J Am Chem Soc.* 2019 Sep. 4; 141(35): 13708-13712. In some embodiments, an intein polypeptide of a fusion protein of the present invention may be configured to be removed (e.g., excised) from the fusion protein and fused with another intein polypeptide (e.g., an intein polypeptide that is a portion of a different fusion protein of the present invention) in situ and/or in vivo.

[0170] In some embodiments, an intein of the present invention is an intein present in a DNA polymerase III gene (DnaE) in cyanobacteria and/or an intein as described in Pinto, F., Thornton, E. L. & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat Commun* 11, 1529 (2020). Further exemplary inteins include, but are not limited, *Nostoc punctiforme* (Npu) inteins and mutants thereof (e.g., NpuGEP, a mutant that contains 3 amino acid residue mutations). In some embodiments, an intein polypeptide is a portion of a *Nostoc punctiforme* (Npu) intein and/or a portion of a mutant Npu intein (e.g., a portion of NpuGEP). An intein polypeptide may be an N-terminal portion of an intein in that the intein polypeptide includes the N-terminus of the full-length intein and/or active complex. In some embodiments, an intein polypeptide may be a C-terminal portion of an intein in that the intein polypeptide includes the C-terminus of the full-length intein and/or active complex. An intein polypeptide that is an N-terminal portion of an intein and an intein polypeptide that contains the remaining portion of the intein (e.g., the C-terminal portion of the intein) together are an intein pair and form the intein and/or the active complex. An intein polypeptide of the present invention may have at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to one or more of SEQ ID NOs: 110-112.

[0171] A Cas12a polypeptide of the present invention may be a portion of a Cas12a protein, optionally a portion of a Cas12a fusion protein. In some embodiments, a Cas12a protein and/or Cas12a fusion protein may be a protein as described in U.S. Patent Application Publication No. 2022/0112473, the contents of which are incorporated herein by reference in its entirety. In some embodiments, a Cas12a polypeptide is portion of a sequence of SEQ ID NO:38-60, 113-149, 192-195, or 196-259. In some embodiments, a Cas12a polypeptide is about 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or more consecutive amino acids of the sequence of SEQ ID NO:38-60, 113-149, 192-195, or 196-259. In some embodiments, a Cas12a polypeptide is one part of two total parts that together form the Cas12a protein. For example, a first Cas12a polypeptide that is present in a first fusion protein of the present invention together with a second Cas12a polypeptide that is present in a second fusion protein of the present invention together form a Cas12a protein. In some embodiments, a Cas12a polypeptide is a N-terminal portion of a Cas12a protein in that the Cas12a polypeptide includes the N-terminus of the full-length Cas12a protein. In some embodiments, a Cas12a polypeptide is a C-terminal portion of a Cas12a protein in that the Cas12a polypeptide includes the C-terminus of the full-length Cas12a protein. Two Cas12a polypeptides that are individually present in two different fusion proteins of the present invention may, upon fusing together of the two Cas12a polypeptides, provide a Cas12a protein that is part of an editing system as described herein such as a CRISPR-Cas editing system. The editing system may be used to modify a target nucleic acid. In some embodiments, a fusion protein of the present invention comprises a Cas12a polypeptide that is a N-terminal portion of a Cas12a protein and an intein polypeptide that is a N-terminal portion of an intein, wherein the intein polypeptide is at the C-terminus of the fusion protein and/or at the C-terminus of the Cas-12a polypeptide. In some embodiments, a fusion protein of the present invention comprises a Cas12a polypeptide that is a C-terminal portion of a Cas12a protein and an intein polypeptide that is a C-terminal portion of an intein, wherein the intein polypeptide is at the N-terminus of the fusion protein and/or at the N-terminus of the Cas-12a polypeptide.

[0172] In some embodiments, a Cas12a protein is split into two portions and a fusion protein of the present invention comprises one of the portions. For example, a Cas12a protein may be split into two portions between amino acid residues 173 and 174, 174 and 175, 175 and 176, 309 and 310, 310 and 311, 405 and 406, 406 and 407, 440 and 441,441 and 442, 549 and 550, or 550 and 551 such that a Cas12a polypeptide includes 173, 174, 175, 309, 310, 405, 406, 440, 441, 549, or 550 consecutive amino acids of a Cas12a protein and another Cas12a polypeptide includes the remaining portion of the Cas12a protein (e.g., from amino acid residue 174, 175, 176, 310, 311, 406, 407, 441, 442, 549, or 551 to the end of the protein). In some embodiments, a Cas12a polypeptide has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to one or more of SEQ ID NOs:150-159 and 175-184.

[0173] In some embodiments, a fusion protein of the present invention comprises all or a portion of a polypeptide of interest. In some embodiments, a polypeptide of interest is fused (directly or via a linker) to a Cas12a polypeptide and/or to an intein polypeptide. In some embodiments, a polypeptide of interest is fused (directly or via a linker) to a Cas12a polypeptide and intein polypeptide such that the polypeptide of interest is between the Cas12a polypeptide and the intein polypeptide. In some embodiments, a polypeptide of interest is fused (directly or via a linker) to a Cas12a polypeptide and intein polypeptide such that the Cas12a polypeptide is between the polypeptide of interest and the intein polypeptide. In some embodiments, a fusion protein of the present invention comprises a polypeptide of interest that is fused (directly or via a linker) to an intein polypeptide and the fusion protein is devoid of a Cas12a polypeptide. In some embodiments, a polypeptide of interest is fused (directly or via a linker) to the N-terminus of an intein polypeptide. In some embodiments, a polypeptide of interest is fused (directly or via a linker) to the C-terminus of an intein polypeptide.

[0174] A fusion protein of the present invention may comprise a reverse transcriptase. A reverse transcriptase polypeptide of the present invention may be all or a portion of a reverse transcriptase. In some embodiments, a reverse transcriptase has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to one or more of SEQ ID NOs:160-171. In some embodiments, a reverse transcriptase is fused (directly or via a linker) to a Cas12a polypeptide such that the Cas12a polypeptide is between the reverse transcriptase and the intein polypeptide. In some embodiments, a reverse transcriptase is fused (directly or via a linker) to a Cas12a polypeptide and intein polypeptide such that the reverse transcriptase is between the Cas12a polypeptide and the intein polypeptide. In some embodiments, a reverse transcriptase is fused (directly or via a linker) to the N-terminus of a Cas12a polypeptide. In some embodiments, a reverse transcriptase is fused (directly or via a linker) to the C-terminus of a Cas12a polypeptide. In some embodiments, a reverse transcriptase polypeptide is fused (directly or via a linker) to an intein polypeptide to provide a fusion protein. In some embodiments, a fusion protein comprising a reverse transcriptase polypeptide and an intein polypeptide is devoid of a Cas12a polypeptide.

[0175] A fusion protein of the present invention may comprise a nuclear localization signal. In some embodiments, a nuclear localization signal has at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to one or more of SEQ ID NOs:172-174.

[0176] In some embodiments, a nucleic acid molecule encoding a fusion protein of the present invention is provided. The nucleic acid molecule may be operably associated with a promoter. In some embodiments, an expression cassette or vector comprising a nucleic acid molecule encoding a fusion protein of the present invention is provided. In some embodiments, an AAV vector comprising a nucleic acid molecule encoding a fusion protein of the present invention is provided.

[0177] A complex comprising a Cas12a protein, a guide nucleic acid (e.g., a guide RNA and/or an extended guide nucleic acid), optionally a reverse transcriptase, and optionally a deaminase may be provided according to embodiments of the present invention. In some embodiments, a complex comprises a Cas12a protein, an extended guide nucleic acid, and a reverse transcriptase. In some embodiments, a complex comprises a Cas12a protein, a guide nucleic acid, and a deaminase. The Cas12a protein of a complex of the present invention may be prepared from a first fusion protein of the present invention and a second fusion protein of the present invention, wherein the first fusion protein comprises a first Cas12a polypeptide fused to a first intein polypeptide and the second fusion protein comprises a second Cas12a polypeptide fused to a second intein polypeptide. Upon contact of the first fusion protein and the second fusion protein (e.g., the first and second fusion proteins being provided together (e.g., in the same composition or cell) under conditions suitable for carrying out the excision of the first and second intein polypeptides, association of the first and second intein polypeptides, and fusion of the first and second Cas12a polypeptides), the first and second intein polypeptides may associate to form an intein (e.g., an active complex), and the intein may excise the intein (e.g., the first and second intein polypeptides) and fuse the first and second Cas12a polypeptides together optionally with a linker (e.g., a peptide linker) between the first and second Cas12a polypeptides.

[0178] In some embodiments, a complex of the present invention comprises an engineered protein (e.g., a fusion protein, a base editor, a templated editor, etc.) and a guide nucleic acid (e.g., a guide RNA). A base editor may comprise a CRISPR-Cas effector protein (e.g., a Cas12a) and a deaminase. In some embodiments, a templated editor may comprise a CRISPR-Cas effector protein (e.g., a Cas12a) and a reverse transcriptase, In some embodiments, a templated editor may be referred to as a REDRAW editor. In some embodiments, a complex of the present invention comprises an engineered protein that is prepared from a first fusion protein of the present invention and a second fusion protein of the present invention, wherein the first fusion protein comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide and the second fusion protein comprises a Cas12a polypeptide fused to a second intein polypeptide; and a guide nucleic acid (e.g., a guide RNA). In some embodiments, a first fusion protein of the present invention and a second fusion protein of the present invention together provide and/or form an engineered protein, wherein the first fusion protein comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide and the second fusion protein comprises a Cas12a polypeptide fused to a second intein polypeptide. In some embodiments, a fusion protein of the present invention comprises, optionally in the N- to C-direction, all or a portion of a polypeptide of interest (e.g., a reverse transcriptase polypeptide), a linker, and an intein polypeptide, optionally wherein the linker comprises a sequence of SEQ ID NO:189. In some embodiments, a fusion protein of the present invention comprises, optionally in the N- to C-direction, an intein polypeptide, a linker, and all or a portion of a Cas12a, optionally wherein the linker comprises a sequence of SEQ ID NO:190.

[0179] In some embodiments, a composition is provided that comprises: a first fusion protein of the present invention that comprises a first Cas12a polypeptide fused to a first intein polypeptide; and a second fusion protein of the present invention that comprises a second Cas12a polypeptide fused to a second intein polypeptide. The first fusion protein and the second fusion protein may be different from each other. In some embodiments, the first intein polypeptide of the first fusion protein and the second intein polypeptide of the second fusion protein together form an intein and/or are two parts that make up a full-length intein and/or the first Cas12a polypeptide of the first fusion protein and the second Cas12a polypeptide of the second fusion protein together form a Cas12a protein and/or are two parts that make up a full-length Cas12a protein. In some embodiments, the first and second fusion proteins are present in the same cell and may optionally be delivered to the cell using separate compositions or a composition comprising both the first and second fusion proteins.

[0180] In some embodiments, a composition of the present invention comprises: a first fusion protein of the present invention that comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide; and a second fusion protein of the present invention that comprises a Cas12a polypeptide fused to a second intein polypeptide. The first fusion protein and the second fusion protein may be different from each other. In some embodiments, the first intein polypeptide of the first fusion protein and the second intein polypeptide of the second fusion protein together form an intein and/or are two parts that make up a full-length intein and/or the polypeptide of interest of the first fusion protein and the Cas12a polypeptide of the second fusion protein together form a fusion protein (e.g., an engineered protein and/or a templated editor) and/or are two parts that make up a full-length fusion protein (e.g., an engineered protein and/or a templated editor). In some embodiments, the first and second fusion proteins are present in the same cell and may optionally be delivered to the cell using separate compositions or a composition comprising both the first and second fusion proteins.

[0181] In some embodiments, a composition of the present invention comprises a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide. The first nucleic acid molecule may encode a fusion protein of the present invention and the second nucleic acid molecule may encode a fusion protein of the present invention. In some embodiments, the first nucleic acid molecule is present in a first expression cassette and/or vector and the second nucleic acid molecule is present in a second expression cassette and/or vector, wherein the first and second expression cassettes and/or vectors are separate from each other and/or are different.

[0182] In some embodiments, a composition of the present invention comprises a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide. The first nucleic acid molecule may encode a fusion protein of the present invention and the second nucleic acid molecule may encode a fusion protein of the present invention. In some embodiments, the first nucleic acid molecule is present in a first expression cassette and/or vector and the second nucleic acid molecule is present in a second expression cassette and/or vector, wherein the first and second expression cassettes and/or vectors are separate from each other and/or are different.

[0183] A kit may be provided according to some embodiments of the present invention. A kit of the present invention may comprise a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide. In some embodiments, a kit of the present invention comprises a first nucleic acid molecule encoding a first fusion protein, the first fusion protein comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide; and a second nucleic acid molecule encoding a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide. In some embodiments, the first nucleic acid molecule of the present invention is present in a first expression cassette and/or vector and the second nucleic acid molecule of the present invention is present in a second expression cassette and/or vector, wherein the first and second expression cassettes and/or vectors are separate from each other and/or are different.

[0184] A complex and/or method of the present invention may use and/or include a Cas12a protein that is provided by (e.g., prepared from) two different fusion proteins of the present invention. For example, when the two different fusion proteins are in contact (with each fusion protein comprising a Cas12a polypeptide), the intein polypeptides of the two fusion proteins may associate to form an intein (e.g., an active complex), and the intein may excise the intein (e.g., the first and second intein polypeptides) and fuse the first and second Cas12a polypeptides together optionally with a linker (e.g., a peptide linker) between the first and second Cas12a polypeptides, to form the Cas12a protein. The two different fusion proteins may be provided by and/or be present in a composition and/or kit of the present invention. In some embodiments, a method of the present invention uses an editing system (e.g., a CRISPR-Cas editing system) in which a Cas12a protein of the present invention (e.g., a Cas12a protein formed by a composition and/or method of the present invention) is part of the editing system and is provided by (e.g., prepared from) two different fusion proteins of the present invention. The editing system may be used to modify a target nucleic acid.

[0185] In some embodiments, a complex and/or method of the present invention may use and/or include a fusion protein (e.g., an engineered protein) that is provided by (e.g., prepared from) two different fusion proteins of the present invention. For example, when a first fusion protein that comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) and a first intein polypeptide is in contact with a second fusion protein that comprises a Cas12a polypeptide and a second intein polypeptide, the intein polypeptides of the two fusion proteins may associate to form an intein (e.g., an active complex), and the intein may excise the intein (e.g., the first and second intein polypeptides) and fuse the polypeptide of interest and Cas12a polypeptide together optionally with a linker (e.g., a peptide linker) between the polypeptide of interest and Cas12a polypeptide, to form a fusion protein (e.g., an engineered protein). The two different fusion proteins may be provided by and/or be present in a composition and/or kit of the present invention. In some embodiments, a method of the present invention uses an editing system (e.g., a CRISPR-Cas editing system) in which a fusion protein of the present invention (e.g., an engineered protein (e.g., a templated editor) formed by a composition and/or method of the present invention) is part of the editing system and is provided by (e.g., prepared from) two different fusion proteins of the present invention. The editing system may be used to modify a target nucleic acid.

[0186] According to some embodiments, provided is a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: a Cas12a protein prepared from a first fusion protein of the present invention and a second fusion protein of the present invention, wherein the first fusion protein comprises a first Cas12a polypeptide fused to a first intein polypeptide and the second fusion protein comprises a second Cas12a polypeptide fused to a second intein polypeptide; and a guide nucleic acid (e.g., a guide RNA). The Cas12a protein and the guide nucleic acid may form a complex or may be comprised in a complex. The target nucleic acid may be present in a cell (e.g., a eukaryotic cell). In some embodiments, the target nucleic acid is present in a plant cell or a human cell. A method of modifying a target nucleic acid may comprise introducing a first nucleic acid molecule encoding the first fusion protein into the cell and introducing a second nucleic acid molecule encoding the second fusion protein into the cell, and expressing the first fusion protein and the second fusion protein in the cell. The first and second nucleic acid molecules may be present in the same composition such that the first and second nucleic acid molecules may be introduced together. In some embodiments, the first and second nucleic acid molecules are in different compositions such that the first and second nucleic acid molecules are introduced together in different, separate compositions or are introduced sequentially in any order. In some embodiments, the first nucleic acid molecule and/or the second nucleic acid molecule is/are present in an expression cassette and/or vector. In some embodiments, the expression cassette and/or vector is an AAV vector. In some embodiments, the target nucleic acid is present in a cell, optionally in a cell of an organism (e.g., a plant, human, etc.). In some embodiments, the method of modifying a target nucleic acid is carried out in vitro, in vivo, or ex vivo.

[0187] In some embodiments, a method of modifying a target nucleic acid of the present invention comprises contacting the target nucleic acid with: an engineered protein (e.g., a templated editor) that is prepared from a first fusion protein and a second fusion protein, wherein the first fusion protein comprises a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide and the second fusion protein comprises a Cas12a polypeptide fused to a second intein polypeptide; and a guide nucleic acid (e.g., a guide RNA). The engineered protein and the guide nucleic acid may form a complex or may be comprised in a complex. The target nucleic acid may be present in a cell (e.g., a eukaryotic cell). In some embodiments, the target nucleic acid is present in a plant cell or a human cell. A method of modifying a target nucleic acid may comprise introducing a first nucleic acid molecule encoding the first fusion protein into the cell and introducing a second nucleic acid molecule encoding the second fusion protein into the cell, and expressing the first fusion protein and the second fusion protein in the cell. The first and second nucleic acid molecules may be present in the same composition such that the first and second nucleic acid molecules may be introduced together. In some embodiments, the first and second nucleic acid molecules are in different compositions such that the first and second nucleic acid molecules are introduced together in different, separate compositions or are introduced sequentially in any order. In some embodiments, the first nucleic acid molecule and/or the second nucleic acid molecule is/are present in an expression cassette and/or vector. In some embodiments, the expression cassette and/or vector is an AAV vector. In some embodiments, the target nucleic acid is present in a cell, optionally in a cell of an organism (e.g., a plant, human, etc.). In some embodiments, the method of modifying a target nucleic acid is carried out in vitro, in vivo, or ex vivo.

[0188] According to some embodiments, provided is a method of modifying a target nucleic acid, the method comprising: introducing a first nucleic acid molecule and a second nucleic acid molecule into a cell, wherein the first nucleic acid molecule encodes a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide and the second nucleic acid molecule encodes a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide; contacting the target nucleic acid in the cell with a protein comprising at least a portion of the first Cas12a polypeptide and at least a portion of the second Cas12a polypeptide and a guide nucleic acid (e.g., a guide RNA and/or an extended guide nucleic acid). In some embodiments, the guide nucleic acid and the protein comprising at least a portion of the first Cas12a polypeptide and at least a portion of the second Cas12a polypeptide form a complex or are comprised in a complex. The method may comprises expressing the first fusion protein and the second fusion protein in the cell. In some embodiments, following the introducing step, the method comprises cleaving (e.g., excising) the first intein polypeptide from the first fusion protein and cleaving (e.g., excising) the second intein polypeptide from the second fusion protein. The method may also comprise prior to, during, and/or after cleaving, associating the first intein polypeptide and the second intein polypeptide to form an intein. The cleaving step may cleave the first Cas12a polypeptide from the first fusion protein and the second Cas12a polypeptide from the second fusion protein and/or the method may further comprise cleaving the first Cas12a polypeptide from the first fusion protein and the second Cas12a polypeptide from the second fusion protein. Prior to, during, and/or after cleaving of the first and second intein polypeptides, the method may comprise fusing the first and second Cas12a polypeptides together (e.g., via a peptide bond between the first and second Cas12a polypeptides) to form a Cas12a protein, wherein the protein that contacts the target nucleic acid in the cell is the Cas12a protein. In some embodiments, the intein fuses the first and second Cas12a polypeptides together. In some embodiments, cleaving of the first intein polypeptide from the first fusion protein and cleaving the second intein polypeptide from the second fusion protein occurs concurrently with fusing the first and second Cas12a polypeptides together. The introducing step may comprise introducing a first expression cassette and/or vector that comprises the first nucleic acid molecule and introducing a second expression cassette and/or vector that comprises the second nucleic acid molecule into the cell. The first and/or second expression cassette and/or vector may comprise the guide nucleic acid, or the method may comprise introducing into the cell a third expression cassette and/or vector that comprises the guide nucleic acid. In some embodiments, a first, second, and/or third expression cassette and/or vector is an AAV vector.

[0189] In some embodiments, a method of modifying a target nucleic acid of the present invention comprises introducing a first nucleic acid molecule and a second nucleic acid molecule into a cell, wherein the first nucleic acid molecule encodes a first fusion protein, the first fusion protein comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) fused to a first intein polypeptide, and the second nucleic acid molecule encodes a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide; contacting the target nucleic acid in the cell with a protein (e.g., a templated editor) comprising at least a portion of the polypeptide of interest and at least a portion of the Cas12a polypeptide and a guide nucleic acid (e.g., a guide RNA). In some embodiments, the guide nucleic acid and the protein comprising at least a portion of the polypeptide of interest and at least a portion of the Cas12a polypeptide form a complex or are comprised in a complex. The method may comprises expressing the first fusion protein and the second fusion protein in the cell. In some embodiments, following the introducing step, the method comprises cleaving (e.g., excising) the first intein polypeptide from the first fusion protein and cleaving (e.g., excising) the second intein polypeptide from the second fusion protein. The method may also comprise prior to, during, and/or after cleaving, associating the first intein polypeptide and the second intein polypeptide to form an intein. The cleaving step may cleave the polypeptide of interest from the first fusion protein and the Cas12a polypeptide from the second fusion protein and/or the method may further comprise cleaving the polypeptide of interest from the first fusion protein and the Cas12a polypeptide from the second fusion protein. Prior to, during, and/or after cleaving of the first and second intein polypeptides, the method may comprise fusing the polypeptide of interest and the Cas12a polypeptide together (e.g., via a peptide bond that is between the polypeptide of interest and the Cas12a polypeptide) to form a fusion protein (optionally wherein the fusion protein is a templated editor), wherein the fusion protein contacts the target nucleic acid in the cell. In some embodiments, the intein fuses the polypeptide of interest and Cas12a polypeptide together. In some embodiments, cleaving of the first intein polypeptide from the first fusion protein and cleaving the second intein polypeptide from the second fusion protein occurs concurrently with fusing the polypeptide of interest and Cas12a polypeptide together. The introducing step may comprise introducing a first expression cassette and/or vector that comprises the first nucleic acid molecule and introducing a second expression cassette and/or vector that comprises the second nucleic acid molecule into the cell. The first and/or second expression cassette and/or vector may comprise the guide nucleic acid, or the method may comprise introducing into the cell a third expression cassette and/or vector that comprises the guide nucleic acid. In some embodiments, a first, second, and/or third expression cassette and/or vector is an AAV vector.

[0190] In some embodiments, a method of the present invention has increased efficiency in modifying a target nucleic acid compared to the efficiency of a control method. An exemplary control method includes a method that contacts a target nucleic acid with a wild-type CRISPR-Cas effector protein that is not fused together via protein splicing (e.g., from two different proteins) optionally using an intein. Another exemplary control method includes a method that contacts a target nucleic acid with a fusion protein (e.g., an engineered protein and/or a templated editor) that is not fused together via protein splicing (e.g., from two different proteins) optionally using an intein. A method of the present invention may generate increased indels and/or increased levels of modification (e.g., precise modifications) compared to a control method. In some embodiments, an editing system used in a method of the present invention is a Redraw editing system such as described in U.S. Patent Application Publication No. 2021/0130835 and/or in U.S. Patent Application Publication No. 2022/0145334, the contents of each of which are incorporated herein by reference in their entirety, but optionally wherein the CRISPR-Cas effector protein is a Cas12a protein that is fused together via protein splicing from two fusion proteins of the present invention and/or wherein the templated editor is a protein that is fused together via protein splicing from two fusion proteins of the present invention.

[0191] According to embodiments of the present invention, a Cas12a protein may be split into two parts (e.g., two Cas12a polypeptides) and each part may individually be fused to a portion (e.g., a fragment) of a trans-splicing split intein (e.g., an intein polypeptide) to provide two different fusion proteins. Each of the two fusion proteins may be separately packaged in an AAV vector. For example, a nucleic acid molecule encoding a fusion portion comprising a Cas12a polypeptide and an intein polypeptide may be provided in an AAV vector. In some embodiments, the two fusions proteins (the Cas12a polypeptide of each form the full-length Cas12a protein and the intein polypeptide of each form the full-length intein) are individually packaged in separate AAV vectors and when both AAV vectors are introduced into (e.g., infect) the same cell, both fusion proteins can be expressed and the two Cas12a polypeptides can be fused (e.g., spliced) together in situ.

[0192] In some embodiments, an engineered protein (e.g., a base editor, templated editor, etc.) may be split into two parts (e.g., one part comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) and a second part comprising a Cas12a polypeptide) and each part may individually be fused to a portion (e.g., a fragment) of a trans-splicing split intein (e.g., an intein polypeptide) to provide two different fusion proteins. Each of the two fusion proteins may be separately packaged in an AAV vector. For example, a nucleic acid molecule encoding a fusion portion comprising a polypeptide of interest (e.g., a reverse transcriptase polypeptide) and an intein polypeptide may be provided in an AAV vector. In some embodiments, the two fusions proteins (of which together form the engineered protein and the intein polypeptides of each form a full-length intein) are individually packaged in separate AAV vectors and when both AAV vectors are introduced into (e.g., infect) the same cell, both fusion proteins can be expressed and the polypeptide of interest and Cas12a polypeptide can be fused (e.g., spliced) together in situ.

[0193] In some embodiments, an editing system of the present invention utilizes the Redraw editing system. Further details on the Redraw editing system can be found in U.S. Patent Application Publication No. 2021/0130835 and/or in U.S. Patent Application Publication No. 2022/0145334, the contents of each of which are incorporated herein by reference in their entirety.

[0194] As described herein, the fusion proteins, nucleic acids, expression cassettes, and/or vectors of the present invention may be codon optimized for expression in an organism. An organism useful with this invention may be any organism or cell thereof for which nucleic acid modification may be useful. An organism can include, but is not limited to, any animal (e.g., a mammal), any plant, any fungus, any archaeon, or any bacterium. In some embodiments, the organism may be a plant or cell thereof. In some embodiments, the organism is an animal such as a mammal (e.g., a human).

[0195] The target nucleic acid may be a genomic sequence from any organism (e.g., eukaryote such as a mammal or a plant). In some embodiments, the target nucleic acid is a genomic sequence from a model organism such as, but not limited to, *Escherichia coli*, an immortalized human cell line (e.g., HEK293, HeLa, etc.), *Caenorhabditis elegans, Arabidopsis thaliana*, and/or *Drosophila Melanogaster*. In some embodiments, the target nucleic acid is a genomic sequence from a non-model organism. Exemplary non-model organisms include, but are not limited to crop plants (e.g., fruit crop plants, vegetable crop plants, and/or field crop plants) and/or animals such as humans, primates and/or mice. In some embodiments, the non-model organism is a crop plant such as corn, soybean, wheat, or canola. In some embodiments, the non-model organism is an animal for testing and/or use of a human therapeutic.

[0196] A target nucleic acid of any plant or plant part may be modified using the nucleic acid constructs of the invention. Any plant (or groupings of plants, for example, into a genus or higher order classification) may be modified using a fusion protein of the invention including an angiosperm, a gymnosperm, a monocot, a dicot, a C3, C4, CAM plant, a bryophyte, a fern and/or fern ally, a microalgae, and/or a macroalgae. A plant and/or plant part useful with this invention may be a plant and/or plant part of any plant species/variety/cultivar. The term “plant part,” as used herein, includes but is not limited to, embryos, pollen, ovules, seeds, leaves, stems, shoots, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, plant cells including plant cells that are intact in plants and/or parts of plants, plant protoplasts, plant tissues, plant cell tissue cultures, plant calli, plant clumps, and the like. As used herein, “shoot” refers to the above ground parts including the leaves and stems. Further, as used herein, “plant cell” refers to a structural and physiological unit of the plant, which comprises a cell wall and also may refer to a protoplast. A plant cell can be in the form of an isolated single cell or can be a cultured cell or can be a part of a higher-organized unit such as, for example, a plant tissue or a plant organ.

[0197] Non-limiting examples of plants useful with the present invention include turf grasses (e.g., bluegrass, bentgrass, ryegrass, fescue), feather reed grass, tufted hair grass, miscanthus, arundo, switchgrass, vegetable crops, including artichokes, kohlrabi, arugula, leeks, asparagus, lettuce (e.g., head, leaf, romaine), malanga, melons (e.g., muskmelon, watermelon, crenshaw, honeydew, cantaloupe), cole crops (e.g., brussels sprouts, cabbage, cauliflower, broccoli, collards, kale, Chinese cabbage, bok choy), cardoni, carrots, napa, okra, onions, celery, parsley, chick peas, parsnips, chicory, peppers, potatoes, cucurbits (e.g., marrow, cucumber, zucchini, squash, pumpkin, honeydew melon, watermelon, cantaloupe), radishes, dry bulb onions, rutabaga, eggplant, salsify, escarole, shallots, endive, garlic, spinach, green onions, squash, greens, beet (sugar beet and fodder beet), sweet potatoes, chard, horseradish, tomatoes, turnips, and spices; a fruit crop such as apples, apricots, cherries, nectarines, peaches, pears, plums, prunes, cherry, quince, fig, nuts (e.g., chestnuts, pecans, pistachios, hazelnuts, pistachios, peanuts, walnuts, macadamia nuts, almonds, and the like), citrus (e.g., clementine, kumquat, orange, grapefruit, tangerine, mandarin, lemon, lime, and the like), blueberries, black raspberries, boysenberries, cranberries, currants, gooseberries, loganberries, raspberries, strawberries, blackberries, grapes (wine and table), avocados, bananas, kiwi, persimmons, pomegranate, pineapple, tropical fruits, pomes, melon, mango, papaya, and lychee, a field crop plant such as clover, alfalfa, timothy, evening primrose, meadow foam, corn/maize (field, sweet, popcorn), hops, jojoba, buckwheat, safflower, quinoa, wheat, rice, barley, rye, millet, sorghum, oats, triticale, sorghum, tobacco, kapok, a leguminous plant (beans (e.g., green and dried), lentils, peas, soybeans), an oil plant (rape, canola, mustard, poppy, olive, sunflower, coconut, castor oil plant, cocoa bean, groundnut, oil palm), duckweed, *Arabidopsis*, a fiber plant (cotton, flax, hemp, jute), *Cannabis* (e.g., *Cannabis sativa, Cannabis indica*, and *Cannabis ruderalis*), lauraceae (cinnamon, camphor), or a plant such as coffee, sugar cane, tea, and natural rubber plants; and/or a bedding plant such as a flowering plant, a cactus, a succulent and/or an ornamental plant (e.g., roses, tulips, violets), as well as trees such as forest trees (broad-leaved trees and evergreens, such as conifers; e.g., elm, ash, oak, maple, fir, spruce, cedar, pine, birch, cypress, eucalyptus, willow), as well as shrubs and other nursery stock. In some embodiments, the nucleic acid constructs of the invention and/or expression cassettes and/or vectors encoding the same may be used to modify maize, soybean, wheat, canola, rice, tomato, pepper, sunflower, raspberry, blackberry, black raspberry and/or cherry.

[0198] In some embodiments, the invention provides cells (e.g., plant cells, animal cells, bacterial cells, archaeon cells, and the like) comprising the polypeptides, polynucleotides, nucleic acid constructs, expression cassettes or vectors of the invention.

[0199] The present invention further comprises a kit or kits to carry out the methods of this invention. A kit of this invention can comprise reagents, buffers, and apparatus for mixing, measuring, sorting, labeling, etc., as well as instructions and the like as would be appropriate for modifying a target nucleic acid.

[0200] In some embodiments, the invention provides a kit for comprising one or more fusion proteins of the present invention, nucleic acid constructs of the present invention, and/or expression cassettes and/or vectors and/or cells comprising the same as described herein, with optional instructions for the use thereof. In some embodiments, a kit may further comprise a CRISPR-Cas guide nucleic acid (corresponding to a Cas12a protein as provided herein, which may be encoded by a polynucleotide of the invention) and/or expression cassettes and/or vectors and or cells comprising the same. In some embodiments, a guide nucleic acid may be provided on the same expression cassette and/or vector as one or more nucleic acid constructs of the invention. In some embodiments, the guide nucleic acid may be provided on a separate expression cassette or vector from that comprising the one or more nucleic acid constructs of the invention.

[0201] Accordingly, in some embodiments, kits are provided comprising a nucleic acid construct comprising (a) a polynucleotide(s) as provided herein and (b) a promoter that drives expression of the polynucleotide(s) of (a). In some embodiments, the kit may further comprise a nucleic acid construct encoding a guide nucleic acid, wherein the construct comprises a cloning site for cloning of a nucleic acid sequence identical or complementary to a target nucleic acid sequence into backbone of the guide nucleic acid.

[0202] In some embodiments, the nucleic acid construct of the invention may be an mRNA that may encode one or more introns within the encoded polynucleotide(s). In some embodiments, the nucleic acid constructs of the invention, and/or an expression cassettes and/or vectors comprising the same, may further encode one or more selectable markers useful for identifying transformants (e.g., a nucleic acid encoding an antibiotic resistance gene, herbicide resistance gene, and the like).

[0203] A polypeptide, polynucleotide, nucleic acid construct, expression cassette, vector, composition, kit, system and/or cell of the present invention may comprise all or a portion of a sequence of one or more of SEQ ID NOs:1-259. In some embodiments, a polypeptide, polynucleotide, nucleic acid construct, expression cassette, vector, composition, kit, system and/or cell of the present invention may comprise at least about 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more consecutive amino acids of a sequence of one or more of SEQ ID NOs:1-259.

[0204] The invention will now be described with reference to the following examples. It should be appreciated that these examples are not intended to limit the scope of the claims to the invention, but are rather intended to be exemplary of certain embodiments. Any variations in the exemplified methods that occur to the skilled artisan are intended to fall within the scope of the invention.

EXAMPLES

Example 1: Validation of Split Intein Protein Reconstitution with mCherry

[0205] The trans-splicing activity of wildtype Npu (SEQ ID NOs:110 and 111, the N- and C-terminal portions, respectively) and NpuGEP (a mutant that contains 3 residue mutations; SEQ ID NOs:110 and 112, the N- and C-terminal portions, respectively) in HEK293T cells was evaluated. For each of the wildtype (WT) Npu and NpuGEP, the N-terminal half of the mCherry protein was fused to the N-terminal portion of the Npu intein (NpuN) and the C-terminal half of the mCherry was fused to the C-terminal portion of the Npu intein (NpuC) (FIG. **1**). After transfecting plasmids encoding two halves of mCherry, their splicing and reconstitution was measured via flow cytometry after 3 days (FIG. **2**). Robust fluorescence was detected from dual plasmid setups approaching the fluorescence intensity of native mCherry (FIG. **2**). Cells where only one half of the mCherry fragments was provided did not show any fluorescence, which demonstrated that both halves of the protein must be present for function (FIG. **2**). This shows that the split intein is functional in human cells and allows rapid splicing of proteins from transient plasmid transfection of each component.

Example 2: Split Intein Protein Reconstitution with Redraw Editors

[0206] Fusion proteins were prepared using a Redraw editor (RE2; SEQ ID NO:113) and the NpuGEP system. Several sites within RE2 were used to introduce a split site (FIG. **3**). In particular, RE2 includes a reverse transcriptase (RT) and Cas12a sequence and fusion proteins were prepared in which a portion of RE2 was included with the split occurring in the Cas12a sequence. Accordingly, some fusion proteins included an N-terminal portion of the Cas12a sequence, which included amino acid residues 1-175, 1-310, 1-406, 1-441, or 1-550 of the Cas12 sequence, and other fusion proteins included the remaining C-terminal portion of the Cas-12a sequence (e.g., amino acid residues 176, 311, 407, 442, or 551 to the end of Cas12a). Fusion proteins having a sequence of SEQ ID NOs:100-109 were prepared. In some cases, two amino acids (cysteine and alanine (CA) or cysteine and phenylalanine (CF)) were inserted between the C-terminal portion of the Npu intein (NpuC) and the C-terminal portion of RE2 (e.g., a C terminal portion of Cas12a), which will leave a "CA" or "CF" scar within the reconstituted RE2 protein. Insertion of the 2 amino acid "scar" residues (here, "CA" or "CF") was confirmed in mature mCherry protein and it was confirmed that the scar did not affect the protein's ability to fluoresce (FIG. **4**).

Example 3: Intein-Mediated Reconstitution of the Redraw Editor in HEK293T Cells

[0207] HEK293T cells were transfected with a plasmid encoding a fusion protein with an N-terminal component of RE2, a plasmid encoding a fusion protein with a C-terminal component of RE2, and a plasmid encoding a stagRNA or crRNA targeting endogenous sites as described in Example 2. After 3 days, high throughput amplicon sequencing was conducted to quantify the Redraw activity at the target sites. It was observed that several split RE2 forms are capable of being spliced together in cotransfected HEK293T cells to enable Redraw activity. Robust indel activity was observed (FIGS. **5**-**6**) as well as precise editing activity (FIG. **7**) that was comparable to the single polypeptide RE2 construct. In addition, it was observed that both halves of RE2 must be expressed in the cell to enable indel and Redraw activity.

[0208] These results demonstrate that the Redraw editor can be separated into two halves (e.g., one half that comprises a reverse transcriptase polypeptide (such as RT(5M)-NpuN) and another half that comprises a Cas12a polypeptide (such as NpuC-Cas12a-Brex27) or e.g., one half that comprises a first Cas12a polypeptide and another half that comprises a second Cas12a polypeptide (such as a RE2 split, for example at 175) and reconstituted in a cell effectively. This can allow the Redraw editor to be separately packaged into viral delivery vehicles (such as adeno associated virus (AAV) vectors) and co-infected to effect Redraw editing in cells targeted by the viral delivery vehicles.

[0209] The foregoing is illustrative of the present invention, 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: 259
SEQ ID NO: 1            moltype = AA  length = 36
FEATURE                 Location/Qualifiers
source                  1..36
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 1
EKSKNDRSKP QPSDDRDRQP PSGEDYPEWK APGEYE                       36

SEQ ID NO: 2            moltype = AA  length = 34
FEATURE                 Location/Qualifiers
source                  1..34
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 2
QEPKPQDQSS EVPPPPGSQK PGTKEPHDSK SSGP                         34

SEQ ID NO: 3            moltype = AA  length = 34
FEATURE                 Location/Qualifiers
source                  1..34
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 3
PDNSSGQKLQ LPQPSDKPQD SREKSDSLPS DKRD                         34

SEQ ID NO: 4            moltype = AA  length = 36
FEATURE                 Location/Qualifiers
source                  1..36
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 4
PDNSTLQTLQ LPQPTPSSTD TQQTSDTDPE DTTDVI                       36

SEQ ID NO: 5            moltype = AA  length = 30
FEATURE                 Location/Qualifiers
source                  1..30
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 5
STSQSDGSSV PADIDQSSDS DQSSSQGQPG                              30

SEQ ID NO: 6            moltype = AA  length = 14
FEATURE                 Location/Qualifiers
source                  1..14
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 6
AKPDDESQKP PQDD                                               14

SEQ ID NO: 7            moltype = AA  length = 14
FEATURE                 Location/Qualifiers
source                  1..14
                        mol_type = protein
```

-continued

```
                       organism = synthetic construct
SEQUENCE: 7
LQLEPGPTTP EYPI                                                      14

SEQ ID NO: 8           moltype = AA  length = 14
FEATURE                Location/Qualifiers
source                 1..14
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 8
IQLPPSDTTP ENPI                                                      14

SEQ ID NO: 9           moltype = AA  length = 12
FEATURE                Location/Qualifiers
source                 1..12
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 9
ESNDNSQVPP SL                                                        12

SEQ ID NO: 10          moltype = AA  length = 10
FEATURE                Location/Qualifiers
source                 1..10
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 10
SEQQEYPGSG                                                           10

SEQ ID NO: 11          moltype = AA  length = 10
FEATURE                Location/Qualifiers
source                 1..10
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 11
NNSEQQENPA                                                           10

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

SEQ ID NO: 13          moltype = AA  length = 20
FEATURE                Location/Qualifiers
source                 1..20
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 13
PKPSSESGER YEQQPEPPPP                                                20

SEQ ID NO: 14          moltype = AA  length = 16
FEATURE                Location/Qualifiers
source                 1..16
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 14
KGGGGEPDEK RPSQSS                                                    16

SEQ ID NO: 15          moltype = AA  length = 14
FEATURE                Location/Qualifiers
source                 1..14
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 15
YAGGTPKEPP PPNS                                                      14

SEQ ID NO: 16          moltype = AA  length = 14
FEATURE                Location/Qualifiers
source                 1..14
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 16
PLVAGGTPFE PPPP                                                      14
```

-continued

```
SEQ ID NO: 17           moltype = AA  length = 14
FEATURE                 Location/Qualifiers
source                  1..14
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 17
PQPDERSQIP DNKE                                                     14

SEQ ID NO: 18           moltype = AA  length = 10
FEATURE                 Location/Qualifiers
source                  1..10
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 18
YTDEKPLPRS                                                          10

SEQ ID NO: 19           moltype = AA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 19
SHPPQEPPQS NL                                                       12

SEQ ID NO: 20           moltype = AA  length = 16
FEATURE                 Location/Qualifiers
source                  1..16
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 20
SESPSKQQPE PKSSKG                                                   16

SEQ ID NO: 21           moltype = AA  length = 16
FEATURE                 Location/Qualifiers
source                  1..16
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 21
SESPTNQQPE PQWTTD                                                   16

SEQ ID NO: 22           moltype = AA  length = 16
FEATURE                 Location/Qualifiers
source                  1..16
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 22
GGSKGPPPSP PPPQPE                                                   16

SEQ ID NO: 23           moltype = AA  length = 16
FEATURE                 Location/Qualifiers
source                  1..16
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 23
GPLPAPPPQP PPPQPN                                                   16

SEQ ID NO: 24           moltype = AA  length = 14
FEATURE                 Location/Qualifiers
source                  1..14
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 24
RPLPHDNNKQ DYSK                                                     14

SEQ ID NO: 25           moltype = AA  length = 61
FEATURE                 Location/Qualifiers
source                  1..61
                        mol_type = protein
                        organism = synthetic construct
VARIANT                 5..61
                        note = Residues are present or absent
SEQUENCE: 25
SGGSGGSGGS GGSGGSGGSG GSGGSGGSGG SGGSGGSGGS GGSGGSGGSG GSGGSGGSGG  60
S                                                                   61

SEQ ID NO: 26           moltype = AA  length = 4
```

-continued

```
FEATURE                Location/Qualifiers
source                 1..4
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 26
SGGS                                                                4

SEQ ID NO: 27          moltype = AA  length = 100
FEATURE                Location/Qualifiers
source                 1..100
                       mol_type = protein
                       organism = synthetic construct
VARIANT                6..100
                       note = Residues are present or absent
SEQUENCE: 27
GGGGSGGGGS GGGGSGGGGS GGGGSGGGGS GGGGSGGGGS GGGGSGGGGS GGGGSGGGGS  60
GGGGSGGGGS GGGGSGGGGS GGGGSGGGGS GGGGSGGGGS                        100

SEQ ID NO: 28          moltype = AA  length = 10
FEATURE                Location/Qualifiers
source                 1..10
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 28
SGGSGGSGGS                                                          10

SEQ ID NO: 29          moltype = AA  length = 16
FEATURE                Location/Qualifiers
source                 1..16
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 29
SGSETPGTSE SATPES                                                   16

SEQ ID NO: 30          moltype = AA  length = 32
FEATURE                Location/Qualifiers
source                 1..32
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 30
SGGSSGGSSG SETPGTSESA TPESSGGSSG GS                                 32

SEQ ID NO: 31          moltype = AA  length = 4
FEATURE                Location/Qualifiers
source                 1..4
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 31
GSSG                                                                4

SEQ ID NO: 32          moltype = AA  length = 6
FEATURE                Location/Qualifiers
source                 1..6
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 32
GSSGSS                                                              6

SEQ ID NO: 33          moltype = AA  length = 8
FEATURE                Location/Qualifiers
source                 1..8
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 33
GSSGSSGS                                                            8

SEQ ID NO: 34          moltype = AA  length = 60
FEATURE                Location/Qualifiers
source                 1..60
                       mol_type = protein
                       organism = synthetic construct
VARIANT                4..60
                       note = Residues are present or absent
SEQUENCE: 34
GSSGSSGSSG SSGSSGSSGS SGSSGSSGSS GSSGSSGSSG SSGSSGSSGS SGSSGSSGSS  60
```

-continued

```
SEQ ID NO: 35          moltype = AA  length = 62
FEATURE                Location/Qualifiers
source                 1..62
                       mol_type = protein
                       organism = synthetic construct
VARIANT                4..60
                       note = Residues may be present or absent
SEQUENCE: 35
GSSGSSGSSG SSGSSGSSGS SGSSGSSGSS GSSGSSGSSG SSGSSGSSGS SGSSGSSGSS  60
GS                                                                 62

SEQ ID NO: 36          moltype = DNA  length = 1592
FEATURE                Location/Qualifiers
source                 1..1592
                       mol_type = other DNA
                       organism = Medicago truncatula
SEQUENCE: 36
actgttaata atttttaaac gtcagcgcac taaaaaaacg aaaagacgga cacgtgaaaa  60
taaaaaacac acactagttt atgacgcaat actattttac ttatgatttg ggtacattag  120
acaaaaccgt gaaagagatg tatcagctat gaaacctgta tacttcaata cagagactta  180
ctcatatcgg atacgtacgc acgaagtatc atattaatta ttttaatttt taataaatat  240
tttatcggat acttatgtga tactctacat atacacaagg atatttctaa gatactttat  300
agatacgtat cctagaaaaa catgaagagt aaaaaagtga gacaatgttg taaaaattca  360
ttataaatgt atatgattca attttagata tgcatcagta taattgattc tcgatgaaac  420
acttaaaatt atatttcttg tggaagaacg tagcgagaga ggtgattcag ttagacaaca  480
ttaaataaaa ttaatgttaa gttcttttaa tgatgtttct ctcaatatca catcatatga  540
aaatgtaata tgatttataa gaaaattttt aaaaaattta ttttaataat cacatgtact  600
atttttaaa aattgtatct tttataataa tacaataata aagagtaatc agtgttaatt  660
tttcttcaaa tataagtttt attataaatc attgttaacg tatcataagt cattaccgta  720
tcgtatctta attttttttt aaaaaccgct aattcacgta cccgtattgt attgtacccg  780
cacctgtatc acaatcgatc ttagttagaa gaattgtctc gaggcggtgc aagacagcat  840
ataatagacg tggactctct tataccaaac gttgtcgtat cacaaagggt taggtaacaa  900
gtcacagttt gtccacgtgt cacgttttaa ttggaagagg tgccgttggc gtaatataac  960
agccaatcga tttttgctat aaaagcaaat caggtaaact aaacttcttc attcttttct  1020
tccccatcgc tacaaaaccg gttcctttgg aaaagagatt cattcaaacc tagcacccaa  1080
ttccgtttca aggtataatc tactttctat tcttcgatta ttttattatt attagctact  1140
atcgtttaat cgatcttttc ttttgatccg tcaaatttaa attcaattag ggttttgttc  1200
ttttctttca tctgattgaa atccttctga attgaaccgt ttacttgatt ttactgttta  1260
ttgtatgatt taatcctttg tttttcaaag acagtcttta gattgtgatt aggggttcat  1320
ataaattttt agatttggat ttttgtattg tatgattcaa aaaatacgtc ctttaattag  1380
attagtacat ggatattttt tacccgattt attgattgtc agggagaatt tgatgagcaa  1440
gttttttga tgtctgttgt aaattgaatt gattataatt gctgatctgc tgcttccagt  1500
tttcataacc catattcttt taaccttgtt gtacacacaa tgaaaaattg gtgattgatt  1560
catttgtttt tctttgtttt ggattataca gg                                1592

SEQ ID NO: 37          moltype = DNA  length = 2000
FEATURE                Location/Qualifiers
source                 1..2000
                       mol_type = other DNA
                       organism = Zea mays
SEQUENCE: 37
gtcgtgcccc tctctagaga taaagagcat tgcatgtcta aagtataaaa aattaccaca  60
tattttttg tcacacttat ttgaagtgta gtttatctat ctctatacat atatttaaac  120
ttcactctac aaataatata gtctataata ctaaaataat attagtgttt tagaggatca  180
tataaataaa ctgctagaca tggtctaaag gataattgaa tattttgaca atctacagtt  240
ttatcttttt agtgtgcatg tgatctctct gttttttttg caaatagctt gacctatata  300
atacttcatc cattttatta gtacatccat ttaggattta gggttgatgg tttctataga  360
ctaattttta gtacatccat tttattcttt ttagtctcta aattttttaa aactaaaact  420
ctattttagt tttttattta ataatttaga tataaaatga aataaaataa attgactaca  480
aataaaacaa ataccottta agaaataaaa aaactaagca aacatttttc ttgtttcgag  540
tagataatga caggctgttc aacgccgtcg acgagtctaa cggacaccaa ccagcgaacc  600
agcagcgtcg cgtcgggcca agcgaagcag acggcacggc atctctgtag ctgcctctgg  660
acccctctcg agagttccgc tccaccgttg gacttgctcc gctgtcggca tccagaaatt  720
gcgtggcgga gcggcagacg tgaggcggca cggcaggcgg cctcttcctc ctctcacggc  780
accggcagct acgggggatt cctttcccac cgctccttcg ctttcccttc ctcgcccgcc  840
gtaataaata gacacccccct ccacaccctc tttccccaac ctcgtgttcg ttcggagcgc  900
acacacacgc aaccagatct cccccaaatc cagccgtcgg cacctccgct tcaaggtacg  960
ccgctcatcc tcccccccccc cctctctcta ccttctctag atcggcgatc cggtccatgg  1020
ttagggcccg gtagttctac ttctgttcat gtttgtgtta gagcaaacat gttcatgttc  1080
atgtttgtga tgatgtggtc tggttgggcg gtcgttctag atcggagtag gatactgttt  1140
caagctacct ggtggattta ttaattttgt atctgtatgt gtgtgccata catcttcata  1200
gttacgagtt taagatgatg gatggaaata tcgatctagg ataggtatac atgttgatgc  1260
gggttttact gatgcatata cagagatgct ttttttctcg cttggttgtg atgatatggt  1320
ctggttgggc ggtcgttcta gatcggagta gaatactgtt tcaaactacc tggtggattt  1380
attaaaggat aaagggtcgt tctagatcgg agtagaatac tgtttcaaac tacctggtgg  1440
atttattaaa ggatctgtat gtatgtgcct acatcttcat agttacgagt ttaagatgat  1500
```

-continued

```
ggatggaaat atcgatctag gataggtata catgttgatg cgggttttac tgatgcatat  1560
acagagatgc ttttttcgc ttggttgtga tgatgtggtc tggttgggcg gtcgttctag  1620
atcggagtag aatactgttt caaactacct ggtggattta ttaattttgt atctttatgt  1680
gtgtgccata catcttcata gttacgagtt taagatgatg gatggaaata ttgatctagg  1740
ataggtatac atgttgatgt gggttttact gatgcatata catgatggca tatgcggcat  1800
ctattcatat gctctaacct tgagtaccta tctattataa taaacaagta tgttttataa  1860
ttattttgat cttgatatac ttggatgatg gcatatgcag cagctatatg tggatttttt  1920
agccctgcct tcatacgcta tttatttgct tggtactgtt tcttttgtcc gatgctcacc  1980
ctgttgtttg gtgatacttc                                              2000

SEQ ID NO: 38          moltype = AA  length = 1227
FEATURE                Location/Qualifiers
REGION                 1..1227
                       note = Lachnospiraceae bacterium
source                 1..1227
                       mol_type = protein
                       organism = unidentified
SEQUENCE: 38
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGYTSDEEVL  300
EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK  360
WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ EYADADLSVV EKLKEIIIQK  420
VDEIYKVYGS SEKLFDADFV LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN  480
RDESFYGDFV LAYDILLKVD HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD  540
YRATILRYGS KYYLAIMDKK YAKCLQKIDK DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK  600
WMAYYNPSED IQKIYKNGTF KKGDMFNLND CHKLIDFFKD SISRYPKWSN AYDFNFSETE  660
KYKDIAGFYR EVEEQGYKVS FESASKKEVD KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT  720
MYFKLLFDEN NHGQIRLSGG AELFMRRASL KKEELVVHPA NSPIANKNPD NPKKTTTLSY  780
DVYKDKRFSE DQYELHIPIA INKCPKNIFK INTEVRVLLK HDDNPYVIGI ARGERNLLYI  840
VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD YHSLLDKKEK ERFEARQNWT SIENIKELKA  900
GYISQVVHKI CELVEKYDAV IALEDLNSGF KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK  960
SNPCATGGAL KGYQITNKFE SFKSMSTQNG FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI  1020
ADSKKFISSF DRIMYVPEED LFEFALDYKN FSRTDADYIK KWKLYSYGNR IRIFRNPKKN  1080
NVFDWEEVCL TSAYKELFNK YGINYQQGDI RALLCEQSDK AFYSSFMALM SLMLQMRNSI  1140
TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE NAILPKNADA NGAYNIARKV LWAIGQFKKA  1200
EDEKLDKVKI AISNKEWLEY AQTSVKH                                      1227

SEQ ID NO: 39          moltype = AA  length = 1307
FEATURE                Location/Qualifiers
source                 1..1307
                       mol_type = protein
                       organism = Acidaminococcus sp.
SEQUENCE: 39
MTQFEGFTNL YQVSKTLRFE LIPQGKTLKH IQEQGFIEED KARNDHYKEL KPIIDRIYKT  60
YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHDYF IGRTDNLTDA  120
INKRHAEIYK GLFKAELFNG KVLKQLGTVT TTEHENALLR SFDKFTTYFS GFYENRKNVF  180
SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV  240
FSFPFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDE TAHIIASLPH  300
RFIPLFKQIL SDRNTLSFIL EEFKSDEEVI QSFCKYKTLL RNENVLETAE ALFNELNSID  360
LTHIFISHKK LETISSALCD HWDTLRNALY ERRISELTGK ITKSAKEKVQ RSLKHEDINL  420
QEIISAAGKE LSEAFKQKTS EILSHAHAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL  480
LDWFAVDESN EVDPEFSARL TGIKLEMEPS LSFYNKARNY ATKKPYSVEK FKLNFQMPTL  540
ASGWDVNKEK NNGAILFVKN GLYYLGIMPK QKGRYKALSF EPTEKTSEGF DKMYYDYFPD  600
AAKMIPKCST QLKAVTAHFQ THTTPILLSN NFIEPLEITK EIYDLNNPEK EPKKFQTAYA  660
KKTGDQKGYR EALCKWIDFT RDFLSKYTKT TSIDLSSLRP SSQYKDLGEY YAELNPLLYH  720
ISFQRIAEKE IMDAVETGKL YLFQIYNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK  780
LNGQAELFYR PKSRMKRMAH RLGEKMLNKK LKDQKTPIPD TLYQELYDYV NHRLSHDLSD  840
EARALLPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNYQ AANSPSKFNQ RVNAYLKEHP  900
ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQQFDY QKKLDNREKE RVAARQAWSV  960
VGTIKDLKQG YLSQVIHEIV DLMIHYQAVV VLENLNFGFK SKRTGIAEKA VYQQFEKMLI  1020
DKLNCLVLKD YPAEKVGGVL NPYQLTDQFT SFAKMGTQSG FLFYVPAPYT SKIDPLTGFV  1080
DPFVWKTIKN HESRKHFLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF  1140
EKNETQFDAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKG IVFRDGSNIL  1200
PKLLENDDSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCFD SRFQNPEWPM  1260
DADANGAYHI ALKGQLLLNH LKESKDLKLQ NGISNQDWLA YIQELRN                1307

SEQ ID NO: 40          moltype = AA  length = 1241
FEATURE                Location/Qualifiers
source                 1..1241
                       mol_type = protein
                       organism = Butyrivibrio proteoclasticus
SEQUENCE: 40
MLLYENYTKR NQITKSLRLE LRPQGKTLRN IKELNLLEQD KAIYALLERL KPVIDEGIKD  60
```

-continued

```
IARDTLKNCE LSFEKLYEHF LSGDKKAYAK ESERLKKEIV KTLIKNLPEG IGKISEINSA  120
KYLNGVLYDF IDKTHKDSEE KQNILSDILE TKGYLALFSK FLTSRITTLE QSMPKRVIEN  180
FEIYAANIPK MQDALERGAV SFAIEYESIC SVDYYNQILS QEDIDSYNRL ISGIMDEDGA  240
KEKGINQTIS EKNIKIKSEH LEEKPFRILK QLHKQILEER EKAFTIDHID SDEEVVQVTK  300
EAFEQTKEQW ENIKKINGFY AKDPGDITLF IVVGPNQTHV LSQLIYGEHD RIRLLLEEYE  360
KNTLEVLPRR TKSEDARYDK FVNAVPKKVA KESHTFDGLQ KMTGDDRLFI LYRDELARNY  420
MRIKEAYGTF ERDILKSRRG IKGNRDVQES LVSFYDELTK FRSALRIINS GNDEKADPIF  480
YNTFDGIFEK ANRTYKAENL CRNYVTKSPA DDARIMASCL GTPARLRTHW WNGEENFAIN  540
DVAMIRRGDE YYYFVLTPDV KPVDLKTKDE TDAQIFVQRK GAKSFLGLPK ALFKCILEPY  600
FESPEHKNDK NCVIEEYVSK PLTIDRRAYD IFKNGTFKKT NIGIDGLTEE KFKDDCRYLI  660
DVYKEFIAVY TRYSCFNMSG LKRADEYNDI GEFFSDVDTR LCTMEWIPVS FERINDMVDK  720
KEGLLFLVRS MFLYNRPRKP YERTFIQLFS DSNMEHTSML LNSRAMIQYR AASLPRRVTH  780
KKGSILVALR DSNGEHIPMH IREAIYKMKN NFDISSEDFI MAKAYLAEHD VAIKKANEDI  840
IRNRRYTEDK FFLSLSYTKN ADISARTLDY INDKVEEDTQ DSRMAVIVTR NLKDLTYVAV  900
VDEKNNVLEE KSLNEIDGVN YRELLKERTK IKYHDKTRLW QYDVSSKGLK EAYVELAVTQ  960
ISKLATKYNA VVVVESMSST FKDKFSFLDE QIFKAFEARL CARMSDLSFN TIKEGEAGSI  1020
SNPIQVSNNN GNSYQDGVIY FLNNAYTRTL CPDTGFVDVF DKTRLITMQS KRQFFAKMKD  1080
IRIDDGEMLF TFNLEEYPTK RLLDRKEWTV KIAGDGSYFD KDKGEYVYVN DIVREQIIPA  1140
LLEDKAVFDG NMAEKFLDKT AISGKSVELI YKWFANALYG IITKKDGEKI YRSPITGTEI  1200
DVSKNTTYNF GKKFMFKQEY RGDGDFLDAF LNYMQAQDIA V                      1241

SEQ ID NO: 41          moltype = AA  length = 1238
FEATURE                Location/Qualifiers
source                 1..1238
                       mol_type = protein
                       organism = Methanoplasma termitum
SEQUENCE: 41
MNNYDEFTKL YPIQKTIRFE LKPQGRTMEH LETFNFFEED RDRAEKYKIL KEAIDEYHKK  60
FIDEHLTNMS LDWNSLKQIS EKYYKSREEK DKKVFLSEQK RMRQEIVSEF KKDDRFKDLF  120
SKKLFSELLK EEIYKKGNHQ EIDALKSFDK FSGYFIGLHE NRKNMYSDGD EITAISNRIV  180
NENFPKFLDN LQKYQEARKK YPEWIIKAES ALVAHNIKMD IVFSLEYFNK VLNQEGIQRY  240
NLALGGYVTK SGEKMMGLND ALNLAHQSEK SSKGRIHMTP LFKQILSEKE SFSYIPDVFT  300
EDSQLLPSIG GFFAQIENDK DGNIFDRALE LISSYAEYDT ERIYIRQADI NRVSNVIFGE  360
WGTLGGLMRE YKADSINDIN LERTCKKVDK WLDSKEFALS DVLEAIDRTG NNDAFNEYIS  420
KMRTAREKID AARKEMKFIS EKISGDEESI HIIKTLLDSV QQFLHFFNLF KARQDIPLDG  480
AFYAEFDEVH SKLFAIVPLY NKVRNYLTKN NLNTKKIKLN FKNPTLANGW DQNKVYDYAS  540
LIFLRDGNYY LGIINPKRKK NIKFEQGSGN GPFYRKMVYK QIPGPNKNLR PVFLTSTKGK  600
KEYKPSKEII EGYEADKHIR GDKFDLDFCH KLIDFFKESI EKHKDWSKFN FYFSPTESYG  660
DISEFYLDVE KQGYRMHFEN ISAETIDEYV EKGDLFLFQI YNKDFVKAAT GKKDMHTIYW  720
NAAFSPENLQ DVVVKLNGEA ELFYRDKSDI KEIVHREGEI LVNRTYNGRT PVPDKIHKKL  780
TDYHNGRTKD LGEAKEYLDK VRYFKAHYDI TKDRRYLNDK IYFHVPLTLN FKANGKKNLN  840
KMVIEKFLSD EKAHIIGIDR GERNLLYYSI IDRSGKIIDQ QSLNVIDGFD YREKLNQREI  900
EMKDARQSWN AIGKIKDLKE GYLSKAVHEI TKMAIQYNAI VVMEELNYGF KRGRFKVEKQ  960
IYQKFENMLI DKMNYLVFKD APDESPGGVL NAYQLTNPLE SFAKLGKQTG ILFYVPAAYT  1020
SKIDPTTGFV NLFNTSSKTN AQERKEFLQK FESISYSAKD GGIFAFAFDY RKFGTSKTDH  1080
KNVWTAYTNG ERMRYIKEKK RNELFDPSKE IKEALTSSGI KYDGGQNILP DILRSNNNGL  1140
IYTMYSSFIA AIQMRVYDGK EDYIISPIKN SKGEFFRTDP KRRELPIDAD ANGAYNIALR  1200
GELTMRAIAE KFDPDSEKMA KLELKHKDWF EFMQTRGD                         1238

SEQ ID NO: 42          moltype = AA  length = 1281
FEATURE                Location/Qualifiers
source                 1..1281
                       mol_type = protein
                       organism = Eubacterium eligens
SEQUENCE: 42
MNGNRSIVYR EFVGVIPVAK TLRNELRPVG HTQEHIIQNG LIQEDELRQE KSTELKNIMD  60
DYYREYIDKS LSGVTDLDFT LLFELMNLVQ SSPSKDNKKA LEKEQSKMRE QICTHLQSDS  120
NYKNIFNAKL LKEILPDFIK NYNQYDVKDK AGKLETLALF NGFSTYFTDF FEKRKNVFTK  180
EAVSTSIAYR IVHENSLIFL ANMTSYKKIS EKALDEIEVI EKNNQDKMGD WELNQIFNPD  240
FYNMVLIQSG IDFYNEICGV VNAHMNLYCQ QTKNNYNLFK MRKLHKQILA YTSTSFEVPK  300
MFEDDMSVYN AVNAFIDETE KGNIIGKLKD IVNKYDELDE KRIYISKDFY ETLSCFMSGN  360
WNLITGCVEN FYDENIHAKG KSKEEKVKKA VKEDKYKSIN DVNDLVEKYI DEKERNEFKN  420
SNAKQYIREI SNIITDTETA HLEYDDHISL IESEEKADEM KKRLDMYMNM YHWAKAFIVD  480
EVLDRDEMFY SDIDDIYNIL ENIVPLYNRV RNYVTQKPYN SKKIKLNFQS PTLANGWSQS  540
KEFDNNAIIL IRDNKYYLAI FNAKNKPDKK IIQGNSDKKN DNDYKKMVYN LLPGANKMLP  600
KVFLSKKGIE TFKPSDYIIS GYNAHKHIKT SENFDISFCR DLIDYFKNSI EKHAEWRKYE  660
FKFSATDSYS DISEFYREVE MQGYRIDWTY ISEADINKLD EEGKIYLFQI YNKDFAENST  720
GKENLHTMYF KNIFSEENLD KIIKLNGQAE LFYRRASVKN PVKHKKDSVL VNKTYKNQLD  780
NGDVVRIPIP DDIYNEIYKM YNGYIKESDL SEAAKEYLDK VEVRTAQKDI VKDYRYTVDK  840
YFIHTPITIN YKVTARNNVN DMVVKYIAQN DDIHVIGIDR GERNLIYISV IDSHGNIVKQ  900
KSYNILNNYD YKKKLVEKEK TREYARKNWK SIGNIKELKE GYISGVVHEI AMLIVEYNAI  960
IAMEDLNYGF KRGRFKVERQ VYQKFESMLI NKLNYFASKE KSVDEPGGLL KGYQLTYVPD  1020
NIKNLGKQCG VIFYVPAAFT SKIDPSTGFI SAFNFKSIST NASRKQFFMQ FDEIRYCAEK  1080
DMFSFGFDYN NFDTYNITMG KTQWTVYTNG ERLQSEFNNA RRTGKTKSIN LTETIKLLLE  1140
DNEINYADGH DIRIDMEKMD EDKKSEFFAQ LLSLYKLTVQ MRNSYTEAEE QENGISYDKI  1200
ISPVINDEGE FFDSDNYKES DDKECKMPKD ADANGAYCIA LKGLYEVLKI KSEWTEDGFD  1260
```

-continued

```
RNCLKLPHAE WLDFIQNKRY E                                         1281

SEQ ID NO: 43          moltype = AA  length = 1300
FEATURE                Location/Qualifiers
source                 1..1300
                       mol_type = protein
                       organism = Francisella tularensis
SEQUENCE: 43
MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF  60
FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK  120
NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK  180
GFHENRKVNY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE  240
ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI  300
NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA  360
AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY  420
ITQQIAPKNL DNPSKKEQEL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA  480
NFAAIPMIFD EIAQNKDNLA QISIKYQNQG KKDLLQASAE DDVKAIKDLL DQTNNLLHKL  540
KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF  600
ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK  660
LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GSPQKGYEKF EFNIEDCRKF  720
IDFYKQSISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ  780
GKLYLFQIYN KDFSAYSKGR PNLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKQSIPKK  840
ITHPAKEAIA NKNKDNPKKE SVFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI  900
NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI  960
EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKLVIEYN AIVVFEDLNF GFKRGRFKVE  1020
KQVYQKLEKM LIEKLNYLVF KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG  1080
FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG  1140
KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAAICGESD  1200
KKFFAKLTSV LNTILQMRNS KTGTELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY  1260
HIGLKGLMLL GRIKNNQEGK KLNLVIKNEE YFEFVQNRNN                      1300

SEQ ID NO: 44          moltype = AA  length = 1206
FEATURE                Location/Qualifiers
REGION                 1..1206
                       note = Lachnospiraceae bacterium
source                 1..1206
                       mol_type = protein
                       organism = unidentified
SEQUENCE: 44
MYYESLTKQY PVSKTIRNEL IPIGKTLDNI RQNNILESDV KRKQNYEHVK GILDEYHKQL  60
INEALDNCTL PSLKIAAEIY LKNQKEVSDR EDFNKTQDLL RKEVVEKLKA HENFTKIGKK  120
DILDLLEKLP SISEDDYNAL ESFRNFYTYF TSYNKVRENL YSDKEKSSTV AYRLINENFP  180
KFLDNVKSYR FVKTAGILAD GLGEEEQDSL FIVETFNKTL TQDGIDTYNS QVGKINSSIN  240
LYNQKNQKAN GFRKIPKMKM LYKQILSDRE ESFIDEFQSD EVLIDNVESY GSVLIESLKS  300
SKVSAFFDAL RESKGKNVYV KNDLAKTAMS VIVFENWRTF DDLLNQEYDL ANENKKKDDK  360
YFEKRQKELK KNKSYSLEHL CNLSEDSCNL IENYIHQISD DIENIIINNE TFLRIVINEH  420
DRSRKLAKNR KAVKAIKDFL DSIKVLEREL KLINSSGQEL EKDLIVYSAH EELLVELKQV  480
DSLYNMTRNY LTKKPFSTEK VKLNFNRSTL LNGWDRNKET DNLGVLLLKD GKYYLGIMNT  540
SANKAFVNPP VAKTEKVFKK VDYKLLPVPN QMLPKVFFAK SNIDFYNPSS EIYSNYKKGT  600
HKKGNMFSLE DCHNLIDFFK ESISKHEDWS KFGFKFDTQA SYNDISEFYR EVEKQGYKLT  660
YTDIDETYIN DLIERNELYL FQIYNKDFSM YSKGKLNLHT LYFMMLFDQR NIDDVVYKLN  720
GEAEVFYRPA SISEDELIIH KAGEEIKNKN PNRARTKETS TFSYDIVKDK RYSKDKFTLH  780
IPITMNFGVD EVKRFNDAVN SAIRIDENVN VIGIDRGERN LLYVVVIDSK GNILEQISLN  840
SIINKEYDIE TDYHALLDER EGGRDKARKD WNTVENIRDL KAGLYLQVVN VVAKLVLKYN  900
AIICLEDLNF GFKRGRQKVE KQVYQKFEKM LIDKLNYLVI DKSREQTSPK ELGGALNALQ  960
LTSKFKSFKE LGKQSGVIYY VPAYLTSKID PTTGFANLFY MKCENVEKSK RFFDGFDFIR  1020
FNALENVFEF GFDYRSFTQR ACGINSKWTV CTNGERIIKY RNPDKNNMFD EKVVVVTDEM  1080
KNLFEQYKIP YEDGRNVKDM IISNEEAEFY RRLYRLLQQT LQMRNSTSDG TRDYIISPVK  1140
NKREAYFNSE LSDGSVPKDA DANGAYNIAR KGLWVLEQIR QKSEGEKINL AMTNAEWLEY  1200
AQTHLL                                                          1206

SEQ ID NO: 45          moltype = AA  length = 1233
FEATURE                Location/Qualifiers
REGION                 1..1233
                       note = Lachnospiraceae bacterium
source                 1..1233
                       mol_type = protein
                       organism = unidentified
SEQUENCE: 45
MDYGNGQFER RAPLTKTITL RLKPIGETRE TIREQKLLEQ DAAFRKLVET VTPIVDDCIR  60
KIADNALCHF GTEYDFSCLG NAISKNDSKA IKKETEKVEK LLAKVLTENL PDGLRKVNDI  120
NSAAFIQDTL TSFVQDDADK RVLIQELKGK TVLMQRFLTT RITALTVWLP DRVFENFNIF  180
IENAEKMRIL LDSPLNEKIM KFDPDAEQYA SLEFYGQCLS QKDIDSYNLI ISGIYADDEV  240
KNPGINEIVK EYNQQIRGDK DESPLPKLKK LHKQILMPVE KAFFVRVLSN DSDARSILEK  300
ILKDTEMLPS KIIEAMKEAD AGDIAVYGSR LHELSHVIYG DHGKLSQIIY DKESKRISEL  360
METLSPKERK ESKKRLEGLE EHIRKSTYTF DELNRYAEKN VMAAYIAAVE ESCAEIMRKE  420
```

-continued

```
KDLRTLLSKE DVKIRGNRHN TLIVKNYFNA WTVFRNLIRI LRRKSEAEID SDFYDVLDDS  480
VEVLSLTYKG ENLCRSYITK KIGSDLKPEI ATYGSALRPN SRWWSPGEKF NVKFHTIVRR  540
DGRLYYFILP KGAKPVELED MDGDIECLQM RKIPNPTIFL PKLVFKDPEA FFRDNPEADE  600
FVFLSGMKAP VTITRETYEA YRYKLYTVGK LRDGEVSEEE YKRALLQVLT AYKEFLENRM  660
IYADLNFGFK DLEEYKDSSE FIKQVETHNT FMCWAKVSSS QLDDLVKSGN GLLFEIWSER  720
LESYYKYGNE KVLRGYEGVL LSILKDENLV SMRTLLNSRP MLVYRPKESS KPMVVHRDGS  780
RVVDRFDKDG KYIPPEVHDE LYRFFNNLLI KEKLGEKARK ILDNKKVKVK VLESERVKWS  840
KFYDEQFAVT FSVKKNADCL DTTKDLNAEV MEQYSESNRL ILIRNTTDIL YYLVLDKNGK  900
VLKQRSLNII NDGARDVDWK ERFRQVTKDR NEGYNEWDYS RTSNDLKEVY LNYALKEIAE  960
AVIEYNAILI IEKMSNAFKD KYSFLDDVTF KGFETKKLAK LSDLHFRGIK DGEPCSFTNP  1020
LQLCQNDSNK ILQDGVIFMV PNSMTRSLDP DTGFIFAIND HNIRTKKAKL NFLSKFDQLK  1080
VSSEGCLIMK YSGDSLPTHN TDNRVWNCCC NHPITNYDRE TKKVEFIEEP VEELSRVLEE  1140
NGIETDTELN KLNERENVPG KVVDAIYSLV LNYLRGTVSG VAGQRAVYYS PVTGKKYDIS  1200
FIQAMNLNRK CDYYRIGSKE RGEWTDFVAQ LIN                              1233

SEQ ID NO: 46          moltype = AA  length = 1227
FEATURE                Location/Qualifiers
REGION                 1..1227
                       note = Lachnospiraceae bacterium
source                 1..1227
                       mol_type = protein
                       organism = unidentified
SEQUENCE: 46
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  300
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  420
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  480
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  540
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  600
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  720
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS  780
YDVYKDKRFS EDQYELHIPI ANINKCPKNI FKINTEVRVL LKHDDNPYVI GIDRGERNLL  840
YIVVVDGKGN IVEQYSLNEI INNFNGIRIK TDYHSLLDKK EKERFEARQN WTSIENIKEL  900
KAGYISQVVH KICELVEKYD AVIALEDLNS GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD  960
KKSNPCATGG ALKGYQITNK FESFKSMSTQ NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT  1020
SIADKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  1080
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS  1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1200
AEDEKLDKVK IASNKEWLEY AQTSVKH                                     1227

SEQ ID NO: 47          moltype = AA  length = 1264
FEATURE                Location/Qualifiers
source                 1..1264
                       mol_type = protein
                       organism = Leptospira inadai
SEQUENCE: 47
MEDYSGFVNI YSIQKTLRFE LKPVGKTLEH IEKKGFLKKD KIRAEDYKAV KKIIDKYHRA  60
YIEEVFDSVL HQKKKKDKTR FSTQFIKEIK EFSELYYKTE KNIPDKERLE ALSEKLRKML  120
VGAFKGEFSE EVAEKYNKNL FSKELIRNEI EKFCETDEER KQVSNFKSFT TYFTGFHSNR  180
QNIYSDEKKS TAIGYRIIHQ NLPKFLDNLK IIESIQRRFK DFPWSDLKKN LKKIDKNIKL  240
TEYFSIDGFV NVLNQKGIDA YNTILGGKSE ESGEKIQGLN EYINLYRQKN NIDRKNPLNV  300
KILFKQILGD RETKSFIPEA FPDDQSVLNS ITEFAKYLKL DKKKKSIIAE LKKFLSSFNR  360
YELDGIYLAN DNSLASISTF LFDDWSFIKK SVSFKYDESV GDPKKKIKSP LKYEKEKEKW  420
LKQKYYTISF LNDAIESYSK SQDEKRVKIR LEAYFAEFKS KDDAKKQFDL LERIEEAYAI  480
VEPLLGAEYP RDRNLKADKK EVGKIKDFLD SIKSLQFFLK PLLSAEIFDE KDLGFYNQLE  540
GYYEEIDISG HLYNKVRNYL TGKIYSKEKF KLNFENSTLL KGWDENREVA NLCVIFREDQ  600
KYYLGVMDKE NNTILSDIPK VKPNELFYEK MVYKLIPTPH MQLPRIIFSS DNLSIYNPSK  660
SILKIREAKS FKEGKNFKLK DCHKFIDFYK ESISKNEDWS RFDFKFSKTS SYENISEFYR  720
EVERQGYNLD FKKVSKFYID SLVEDGKLYL FQIYNKDFSI FSKGKPNLHT IYFRSLFSKE  780
NLKDVCLKLN GEAEMFFRKK SINYDEKKKR EGHHPELFEK LKYPILKDKR YSEDKFQFHL  840
PISLNFKSKE RLNFNLKVNE FLKRNKDINI IGIDRGERNL LYLVMINQKG EILKQTLLDS  900
MQSGKGRPEI NYKEKLQEKE IERDKARKSW GTVENIKELK EGYLSIVIHQ ISKLMVENNA  960
IVVLEDLNIG FKRGRQKVER QVYQKFEKML IDKLNFLVFK ENKPTEPGGV LKAYQLTDEF  1020
QSFEKLSKQT GFLFYVPSWN TSKIDPRTGF IDFLHPAYEN IEKAKQWINK FDSIRFNSKM  1080
DWFEFTADTR KFSENLMLGK NRVWVICTTN VERYFTSKTA NSSIQYNSIQ ITEKLKELFV  1140
DIPFSNGQDL KPEILRKNDA VFFKSLLFYI KTTLSLRQNN GKKGEEEKDF ILSPVVDSKG  1200
RFFNSLEASD DEPKDADANG AYHIALKGLM NLLVLNETKE ENLSRPKWKI KNKDWLEFVW  1260
ERNR                                                              1264

SEQ ID NO: 48          moltype = AA  length = 1373
FEATURE                Location/Qualifiers
```

-continued

```
source                  1..1373
                        mol_type = protein
                        organism = Moraxella bovoculi
SEQUENCE: 48
MLFQDFTHLY PLSKTVRFEL FIDRTLEHIH AKNFLSQDET MADMHQKVKV ILDDYHRDFI  60
ADMMGEVKLT KLAEFYDVYL KFRKNPKDDE LQKAQLKDLQ AVLRKEIVKP IGNGGKYKAG  120
YDRLFGAKLF KDGKELGDLA KFVIAQEGES SPKLAHLAHF EKFSTYFTGF HDNRKNMYSD  180
EDKHTAIAYR LIHENLPRFI DNLQILTTIK QKHSALYDQI INELTASGLD VSLASHLDGY  240
HKLLTQEGIT AYNTLLGGIS GEAGSPKIQG INELINSHHN QHCHKSERIA KLRPLHKQIL  300
SDGMSVSFLP SKFADDSEMC QAVNEFYRHY ADVFAKVQSL FDGFDDHQKD GIYVEHKNLN  360
ELSKQAFGDF ALLGRVLDGY YVDVVNPEFN ERFAKAKTDN AKAKLTKEKD KFIKGVHSLA  420
SLEQAIEHYT ARHDDESVQA GKLGQYFKHG LAGVDNPIQK IHNNHSTIKG FLERERPAGE  480
RALPKIKSGK NPEMTQLRQL KELLDNALNV AHFAKLLTTK TTLDNQDGNF YGEFGVLYDE  540
LAKIPTLYNK VRDYLSQKPF STEKYKLNFG NPTLLNGWDL NKEKDNFGVI LQKDGCYYLA  600
LLDKAHKKVF DNAPNTGKSI YQKMIYKYLE VRKQFPKVFF SKEAIAINYH PSKELVEIKD  660
KGRQRSDDER LKLYRFILEC LKIHPKYDKK FEGAIGDIQL FKKDKKGREV PISEKDLFKD  720
INGIFSSKPK LEMEDFFIGE FKRYNPSQDL VDQYNIYKKI DSNDNRKKEN FYNNHPKFKK  780
DLVRYYYESM CKHEEWEESF EFSKKLQDIG CYVDVNELFT EIETRRLNYK ISFCNINADY  840
IDELVEQGQL YLFQIYNKDF SPKAHGKPNL HTLYFKALFS EDNLADPIYK LNGEAQIFYR  900
KASLDMNETT IHRAGEVLEN KNPDNPKKRQ FVYDIIKDKR YTQKDFMLHV PITMNFGVQG  960
MTIKEFNKKV NQSIQQYDEV NVIGIDRGER HLLYLTVINS KGEILEQCSL NDITTASANG  1020
TQMTTPYHKI LDKREIERLN ARVGWGEIET IKELKSGYLS HVVHQISQLM LKYNAIVVLE  1080
DLNFGFKRGR FKVEKQIYQN FENALIKKLN HLVLKDKADD EIGSYKNALQ LTNNFTDLKS  1140
IGKQTGFLFY VPAWNTSKID PETGFVDLLK PRYENIQASQ AFFGKFDKIC YNADKDYFEF  1200
HIDYAKFTDK AKNSRQIWTI CSHGDKRYVY DKTANQNKGA AKGINVNDIL KSLFARHHIN  1260
EKQPNLVMDI CQNNDKEFHK SLMYLLKTLL ALRYSNASSD EDFILSPVAN DEGVFFNSAL  1320
ADDTQPQNAD ANGAYHIALK GLWLLNELKN SDDLNKVKLA IDNQTWLNFA QNR         1373

SEQ ID NO: 49           moltype = AA  length = 1352
FEATURE                 Location/Qualifiers
REGION                  1..1352
                        note = Parcubacteria bacterium
source                  1..1352
                        mol_type = protein
                        organism = unidentified
SEQUENCE: 49
MENIFDQFIG KYSLSKTLRF ELKPVGKTED FLKINKVFEK DQTIDDSYNQ AKFYFDSLHQ  60
KFIDAALASD KTSELSFQNF ADVLEKQNKI ILDKKREMGA LRKRDKNAVG IDRLQKEIND  120
AEDIIQKEKE KIYKDVRTLF DNEAESWKTY YQEREVDGKK ITESKADLKQ KGADFLTAAG  180
ILKVLKYEFP EEKEKEFQAK NQPSLFVEEK ENPGQKRYIF DSFDKFAGYL TKFQQTKKNL  240
YAADGTSTAV ATRIADNFII FHQNTKVFRD KYKNNHTDLG FDEENIFEIE RYKNCLLQRE  300
IEHIKNENSY NKIIGRINKK IKEYRDQKAK DTKLTKSDFP FFKNLDKQIL GEVEKEKQLI  360
EKTREKTEED VLIERFKEFI ENNEERFTAA KKLMNAFCNG EFESEYEGIY LKNKAINTIS  420
RRWFVSDRDF ELKLPQQKSK NKSEKNEPKV KKFISIAEIK NAVEELDGDI FKAVFYDKKI  480
IAQGGSKLEQ FLVIWKYEFE YLFRDIEREN GEKLLGYDSC LKIAKQLGIF PQEKEAREKA  540
TAVIKNYADA GLGIFQMMKY FSLDDKDRKN TPGQLSTNFY AEYDGYYKDF EFIKYYNEFR  600
NFITKKPFDE DKIKLNFENG ALLKGWDENK EYDFMGVILK KEGRLYLGIM HKNHRKLFQS  660
MGNAKGDNAN RYQKMIYKQI ADASKDVPRL LLTSKKAMEK FKPSQEILRI KKEKTFKRES  720
KNFSLRDLHA LIEYYRNCIP QYSNWSFYDF QFQDTGKYQN IKEFTDDVQK YGYKISFRDI  780
DDEYINQALN EGKMYLFEVV NKDIYNTKNG SKNLHTLYFE HILSAENLND PVFKLSGMAE  840
IFQRQPSVNE REKITTQKNQ CILDKGDRAY KYRRYTEKKI MFHMSLVLNT GKGEIKQVQF  900
NKIINQRISS SDNEMRVNVI GIDRGEKNLL YYSVVKQNGE IIEQASLNEI NGVNYRDKLI  960
EREKERLKNR QSWKPVVKIK DLKKGYISHV IHKICQLIEK YSAIVVLEDL NMRFKQIRGG  1020
IERSVYQQFE KALIDKLGYL VFKDNRDLRA PGGVLNGYQL SAPFVSFEKM RKQTGILFYT  1080
QAEYTSKTDP ITGFRKNVYI SNSASLDKIK EAVKKFDAIG WDGKEQSYFF KYNPYNLADE  1140
KYKNSTVSKE WAIFASAPRI RRQKGEDGYW KYDRVKVNEE FEKLLKVWNF VNPKATDIKQ  1200
EIIKKIKAGD LQGEKELDGR LRNFWHSFIY LFNLVLELRN SFSLQIKIKA GEVIAVDEGV  1260
DFIASPVKPF FTTPNPYIPS NLCWLAVENA DANGAYNIAR KGVMILKKIR EHAKKDPEFK  1320
KLPNLFISNA EWDEAARDWG KYAGTTALNL DH                                1352

SEQ ID NO: 50           moltype = AA  length = 1260
FEATURE                 Location/Qualifiers
source                  1..1260
                        mol_type = protein
                        organism = Porphyromonas crevioricanis
SEQUENCE: 50
MDSLKDFTNL YPVSKTLRFE LKPVGKTLEN IEKAGILKED EHRAESYRRV KKIIDTYHKV  60
FIDSSLENMA KMGIENEIKA MLQSFCELYK KDHRTEGEDK ALDKIRAVLR GLIVGAFTGV  120
CGRRENTVQN EKYESLFKEK LIKEILPDFV LSTEAESLPF SVEEATRSLK EFDSFTSYFA  180
GFYENRKNIY STKPQSTAIA YRLIHENLPK FIDNILVFQK IKEPIAKELE HIRADFSAGG  240
YIKKDERLED IFSLNYYIHV LSQAGIEKYN ALIGKIVTEG DGEMKGLNEH INLYNQQRGR  300
EDRLPLFRPL YKQILSDREQ LSYLPESFEK DEELLRALKE FYDHIAEDIL GRTQQLMTSI  360
SEYDLSRIYV RNDSQLTDIS KKMLGDWNAI YMARERAYDH EQAPKRITAK YERDRIKALK  420
GEESISLANL NSCIAFLDNV RDCRVDTYLS TLGQKEGPHG LSNLVENVFA SYHEAEQLLS  480
FPYPEENNLI QDKDNVVLIK NLLDNISDLQ RFLKPLWGMG DEPDKDERFY GEYNYIRGAL  540
DQVIPLYNKV RNYLTRKPYS TRKVKLNFGN SQLLSGWDRN KEKDNSCVIL RKGQNFYLAI  600
```

-continued

```
MNNRHKRSFE NKMLPEYKEG EPYFEKMDYK FLPDPNKMLP KVFLSKKGIE IYKPSPKLLE  660
QYGHGTHKKG DTFSMDDLHE LIDFFKHSIE AHEDWKQFGF KFSDTATYEN VSSFYREVED  720
QGYKLSFRKV SESYVYSLID QGKLYLFQIY NKDFSPCSKG TPNLHTLYWR MLFDERNLAD  780
VIYKLDGKAE IFFREKSLKN DHPTHPAGKP IKKKSRQKKG EESLFEYDLV KDRRYTMDKF  840
QFHVPITMNF KCSAGSKVND MVNAHIREAK DMHVIGIDRG ERNLLYICVI DSRGTILDQI  900
SLNTINDIDY HDLLESRDKD RQQEHRNWQT IEGIKELKQG YLSQAVHRIA ELMVAYKAVV  960
ALEDLNMGFK RGRQKVESSV YQQFEKQLID KLNYLVDKKK RPEDIGGLLR AYQFTAPFKS  1020
FKEMGKQNGF LFYIPAWNTS NIDPTTGFVN LFHVQYENVD KAKSFFQKFD SISYNPKKDW  1080
FEFAFDYKNF TKKAEGSRSM WILCTHGSRI KNFRNSQKNG QWDSEEFALT EAFKSLFVRY  1140
EIDYTADLKT AIVDEKQKDF FVDLLKLFKL TVQMRNSWKE KDLDYLISPV AGADGRFFDT  1200
REGNKSLPKD ADANGAYNIA LKGLWALRQI RQTSEGGKLK LAISNKEWLQ FVQERSYEKD  1260

SEQ ID NO: 51          moltype = AA  length = 1324
FEATURE                Location/Qualifiers
source                 1..1324
                       mol_type = protein
                       organism = Prevotella disiens
SEQUENCE: 51
MENYQEFTNL FQLNKTLRFE LKPIGKTCEL LEEGKIFASG SFLEKDKVRA DNVSYVKKEI  60
DKKHKIFIEE TLSSFSISND LLKQYFDCYN ELKAFKKDCK SDEEEVKKTA LRNKCTSIQR  120
AMREAISQAF LKSPQKKLLA IKNLIENVFK ADENVQHFSE FTSYFSGFET NRENFYSDEE  180
KSTSIAYRLV HDNLPIFIKN IYIFEKLKEQ FDAKTLSEIF ENYKLYVAGS SLDEVFSLEY  240
FNNTLTQKGI DNYNAVIGKI VKEDKQEIQG LNEHINLYNQ KHKDRRLPFF ISLKKQILSD  300
REALSWLPDM FKNDSEVIDA LKGFYIEDGF ENNVLTPLAT LLSSLDKYNL NGIFIRNNEA  360
LSSLSQNVYR NFSIDEAIDA QNAELQTFNN YELIANALRA KIKKETKQGR KSFEKYEEYI  420
DKKVKAIDSL SIQEINELVE NYVSEFNSNS GNMPRKVEDY FSLMRKGDFG SNDLIENIKT  480
KLSAAEKLLG TKYQETAKDI FKKDENSKLI KELLDATKQF QHFIKPLLGT GEEADRDLVF  540
YGDFLPLYEK FEELTLLYNK VRNRLTQKPY SKDKIRLCFN KPKLMTGWVD SKTEKSDNGT  600
QYGGYLFRKK NEIGEYDYFL GISSKAQLFR KNEAVIGDYE RLDYYQPKAN TIYGSAYEGE  660
NSYKEDKKRL NKVIIAYIEQ IKQTNIKKSI IESISKYPNI SDDDKVTPSS LLEKIKKVSI  720
DSYNGILSFK SFQSVNKEVI DNLLKTISPL KNKAEFLDLI NKDYQIFTEV QAVIDEICKQ  780
KTFIYFPISN VELEKEMGDK DKPLCLFQIS NKDLSFAKTF SANLRKKRGA ENLHTMLFKA  840
LMEGNQDNLD LGSGAIFYRA KSLDGNKPTH PANEAIKCRN VANKDKVSLF TYDIYKNRRY  900
MENKFLFHLS IVQNYKAAND SAQLNSSATE YIRKADDLHI IGIDRGERNL LYYSVIDMKG  960
NIVEQDSLNI IRNNDLETDY HDLLDKREKE RKANRQNWEA VEGIKDLKKG YLSQAVHQIA  1020
QLMLKYNAII ALEDLGQMFV TRGQKIEKAV YQQFEKSLVD KLSYLVDKKR PYNELGGILK  1080
AYQLASSITK NNSDKQNGFL FYVPAWNTSK IDPVTGFTDL LRPKAMTIKE AQDFFGAFDN  1140
ISYNDKGYFE FETNYDKFKI RMKSAQTRWT ICTFGNRIKR KKDKNYWNYE EVELTEEFKK  1200
LFKDSNIDYE NCNLKEEIQN KDNRKFFDDL IKLLQLTLQM RNSDDKGNDY IISPVANAEG  1260
QFFDSRNGDK KLPLDADANG AYNIARKGLW NIRQIKQTKN KDDLNLSISS TEWLDFVREK  1320
PYLK                                                               1324

SEQ ID NO: 52          moltype = AA  length = 1484
FEATURE                Location/Qualifiers
REGION                 1..1484
                       note = Peregrinibacteria bacterium
SITE                   1073
                       note = misc_feature - Xaa can be any naturally occurring
                        amino acid
source                 1..1484
                       mol_type = protein
                       organism = unidentified
SEQUENCE: 52
MSNFFKNFTN LYELSKTLRF ELKPVGDTLT NMKDHLEYDE KLQTFLKDQN IDDAYQALKP  60
QFDEIHEEFI TDSLESKKAK EIDFSEYLDL FQEKKELNDS EKKLRNKIGE TFNKAGEKWK  120
KEKYPQYEWK KGSKIANGAD ILSCQDMLQF IKYKNPEDEK IKNYIDDTLK GFFTYFGGFN  180
QNRANYYETK KEASTAVATR IVHENLPKFC DNVIQFKHII KRKKDGTVEK TERKTEYLNA  240
YQYLKNNNKI TQIKDAETEK MIESTPIAEK IFDVYYFSSC LSQKQIEEYN RIIGHYNLLI  300
NLYNQAKRSE GKHLSANEKK YKDLPKFKTL YKQIGCGKKK DLFYTIKCDT EEEANKSRNE  360
GKESHSVEEI INKAQEAINK YFKSNNDCEN INTVPDFINY ILTKENYEGV YWSKAAMNTI  420
SDKYFANYHD LQDRLKEAKV FQKADKKSED DIKIPEAIEL SGLFGVLDSL ADWQTTLFKS  480
SILSNEKLKI ITDSQTPSEA LLKMIFNDIE KNMESFLKET NDIITLKKYK GNKEGTEKIK  540
QWFDYTLAIN RMLKYFLVKE NKIKGNSLDT NISEALKTLI YSDDAEWFKW YDALRNYLTQ  600
KPQDEAKENK LKLNFDNPSL AGGWDVNKEC SNFCVILKDK NEKKYLAMIK KGENTLFQKE  660
WTEGRGKNLT KKSNPLFEIN NCEILSKMEY DFWADVSKMI PKCSTQLKAV VNHFKQSDNE  720
FIFPIGYKVT SGEKFREECK ISKQDFELNN KVFNKNELSV TAMRYDLSST QEKQYIKAFQ  780
KEYWELLFKQ EKRDTKLTNN EIFNEWINFC NKKYSELLSW ERKYKDALTN WINFCKYFLS  840
KYPKTTLFNY SFKESENYNS LDEFYRDVDI CSYKLNINTT INKSILDRLV EEGKLYLFEI  900
KNQDSNDGKS IGHKNNLHTI YWNAIFENFD NRPKLNGEAE IFYRKAISKD KLGIVKGKKT  960
KNGTWIIKNY RFSKEKFILH VPITLNFCSN NEYVNDIVNT KFYNFSNLHF LGIDRGEKHL  1020
AYYSLVNKNG EIVDQGTLNL PFTDKDGNQR SIKKEKYFYN KQEDKWEAKE VDXWNYNDLL  1080
DAMASNRDMA RKNWQRIGTI KEAKNGYVSL VIRKIADLAV NNERPAFIVL EDLNTGFKRS  1140
RQKIDKSVYQ KFELALAKKL NFLVDKNAKR DEIGSPTKAL QLTPPVNNYG DIENKKQAGI  1200
MLYTRANYTS QTDPATGWRK TIYLKAGPEE TTYKKDGKIK NKSVKDQIIE TFTDIGFDGK  1260
DYYFEYDKGE FVDEKTGEIK PKKWRLYSGE NGKSLDRFRG EREKDKYEWK IDKIDIVKIL  1320
DDLFVNFDKN ISLLKQLKEG VELTRNNEHG TGESLRFAIN LIQQIRNTGN NERDNDFILS  1380
```

-continued

```
PVRDENGKHF DSREYWDKET KGEKISMPSS GDANGAFNIA RKGIIMNAHI LANSDSKDLS  1440
LFVSDEEWDL HLNNKTEWKK QLNIFSSRKA MAKRKKKRPA ATKK                   1484

SEQ ID NO: 53          moltype = AA  length = 1245
FEATURE                Location/Qualifiers
source                 1..1245
                       mol_type = protein
                       organism = Porphyromonas macacae
SEQUENCE: 53
MKTQHFFEDF TSLYSLSKTI RFELKPIGKT LENIKKNGLI RRDEQRLDDY EKLKKVIDEY  60
HEDFIANILS SFSFSEEILQ SYIQNLSISE ARAKIEKTMR DTLAKAFSED ERYKSIFKKE  120
LVKKDIPVWC PAYKSLCKKF DNFTTSLVPF HENRKNLYTS NEITASIPYR IVHVNLPKFI  180
QNIEALCELQ KKMGADLYLE MMENLRNVWP SFVKTPDDLC NLKTYNHLMV QSSISEYNRF  240
VGGYSTEDGT KHQGINEWIN IYRQRNKEMR LPGLVFLHKQ ILAKVDSSSF ISDTLENDDQ  300
VFCVLRQFRK LFWNTVSSKE DDAASLKDLF CGLSGYDPEA IYVSDAHLAT ISKNIFDRWN  360
YISDAIRRKT EVLMPRKKES VERYAEKISK QIKKRQSYSL AELDDLLAHY SEESLPAGFS  420
LLSYFTSLGG QKYLVSDGEV ILYEEGSNIW DEVLIAFRDL QVILDKDFTE KKLGKDEEAV  480
SVIKKALDSA LRLRKFFDLL SGTGAEIRRD SSFYALYTDR MDKLKGLLKM YDKVRNYLTK  540
KPYSIEKFKL HFDNPSLLSG WDKNKELNNL SVIFRQNGYY YLGIMTPKGK NLFKTLPKLG  600
AEEMFYEKME YKQIAEPMLM LPKVFFPKKT KPAFAPDQSV VDIYNKKTFK TGQKGFNKKD  660
LYRLIDFYKE ALTVHEWKLF NFSFSPTEQY RNIGEFFDEV REQAYKVSMV NVPASYIDEA  720
VENGKLYLFQ IYNKDFSPYS KGIPNLHTLY WKALFSEQNQ SRVYKLCGGG ELFYRKASLH  780
MQDTTVHPKG ISIHKKNLNK KGETSLFNYD LVKDKRFTED KFFFHVPISI NYKNKKITNV  840
NQMVRDYIAQ NDDLQHGIDR GERNLLYISR IDTRGNLLEQ FSLNVIESDK GDLRTDYQKI  900
LGDREQERLR RRQEWKSIES IKDLKDGYMS QVVHKICNMV VEHKAIVVLE NLNLSFMKGR  960
KKVEKSVYEK FERMLVDKLN YLVVDKKNLS NEPGGLYAAY QLTNPLFSFE ELHRYPQSGI  1020
LFFVDPWNTS LTDPSTGFVN LLGRINYTNV GDARKFFDRF NAIRYDGKGN ILFDLDLSRF  1080
DVRVETQRKL WTLTTFGSRI AKSKKSGKWM VERIENLSLC FLELFEQFNI GYRVEKDLKK  1140
AILSQDRKEF YVRLIYLFNL MMQIRNSDGE EDYILSPALN EKNLQFDSRL IEAKDLPVDA  1200
DANGAYNVAR KGLMVVQRIK RGDHESIHRI GRAQWLRYVQ EGIVE                  1245

SEQ ID NO: 54          moltype = AA  length = 1250
FEATURE                Location/Qualifiers
source                 1..1250
                       mol_type = protein
                       organism = Smithella sp.
SEQUENCE: 54
MQTLFENFTN QYPVSKTLRF ELIPQGKTKD FIEQKGLLKK DEDRAEKYKK VKNIIDEYHK  60
DFIEKSLNGL KLDGLEKYKT LYLKQEKDDK DKKAFDKEKE NLRKQIANAF RNNEKFKTLF  120
AKELIKNDLM SFACEEDKKN VKEFEAFTTY FTGFHQNRAN MYVADEKRTA IASRLIHENL  180
PKFIDNIKIF EKMKKEAPEL LSPFNQTLKD MKDVIKGTTL EEIFSLDYFN KTLTQSGIDI  240
YNSVIGGRTP EEGKTKIKGL NEYINTDFNQ KQTDKKKRQP KFKQLYKQIL SDRQSLSFIA  300
EAFKNDTEIL EAIEKFYVNE LLHFSNEGKS TNVLDAIKNA VSNLESFNLT KMYFRSGASL  360
TDVSRKVFGE WSIINRALDN YYATTYPIKP REKSEKYEER KEKWLKQDFN VSLIQTAIDE  420
YDNETVKGKN SGKVIADYFA KFCDDKETDL IQKVNEGYIA VKDLLNTPCP ENEKLGSNKD  480
QVKQIKAFMD SIMDIMHFVR PLSLKDTDKE KDETFYSLFT PLYDHLTQTI ALYNKVRNYL  540
TQKPYSTEKI KLNFENSTLL GGWDLNKETD NTAIILRKDN LYYLGIMDKR HNRIFRNVPK  600
ADKKDFCYEK MVYKLLPGAN KMLPKVFFSQ SRIQEFTPSA KLLENYANET HKKGDNFNLN  660
HCHKLIDFFK DSINKHEDWK NFDFRFSATS TYADLSGFYH EVEHQGYKIS FQSVADSFID  720
DLVNEGKLYL FQIYNKDFSP FSKGKPNLHT LYWKMLFDEN NLKDVVYKLN GEAEVFYRKK  780
SIAEKNTTIH KANESIINKN PDNPKATSTF NYDIVKDKRY TIDKFQFHIP ITMNFKAEGI  840
FNMNQRVNQF LKANPDINII GIDRGERHLL YYALINQKGK ILKQDTLNVI ANEKQKVDYH  900
NLLDKKEGDR ATARQEWGVI ETIKELKEGY LSQVIHKLTD LMIENNAIIV MEDLNFGFKR  960
GRQKVEKQVY QKFEKMLIDK LNYLVDKNKK ANELGGLLNA FQLANKFESF QKMGKQNGFI  1020
FYVPAWNTSK TDPATGFIDF LKPRYENLNQ AKDFFEKFDS IRLNSKADYF EFAFDFKNFT  1080
EKADGGRTKW TVCTTNEDRY QWNRALNNNR GSQEKYDITA ELKSLFDGKV DYKSGKDLKQ  1140
QIASQESADF FKALMKNLSI TLSLRHNNGE KGDNEQDYIL SPVADSKGRF FDSRKADDDM  1200
PKNADANGAY HIALKGLWCL EQISKTDDLK KVKLAISNKE WLEFVQTLKG             1250

SEQ ID NO: 55          moltype = AA  length = 1228
FEATURE                Location/Qualifiers
REGION                 1..1228
                       note = Lachnospiraceae bacterium
source                 1..1228
                       mol_type = protein
                       organism = unidentified
SEQUENCE: 55
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  300
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  420
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  480
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  540
```

-continued

```
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  600
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  720
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS  780
YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY  840
IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK  900
AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK  960
KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS  1020
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  1080
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS  1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1200
AEDEKLDKVK IAISNKEWLE YAQTSVKH                                    1228

SEQ ID NO: 56          moltype = AA  length = 1228
FEATURE                Location/Qualifiers
source                 1..1228
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 56
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  300
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  420
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  480
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  540
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  600
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  720
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS  780
YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY  840
IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK  900
AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK  960
KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS  1020
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  1080
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS  1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1200
AEDEKLDKVK IAISNKEWLE YAQTSVKH                                    1228

SEQ ID NO: 57          moltype = AA  length = 1228
FEATURE                Location/Qualifiers
source                 1..1228
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 57
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  300
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  420
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  480
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  540
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  600
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  720
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVAP ANSPIANKNP DNPKKTTTLS  780
YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY  840
IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK  900
AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK  960
KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS  1020
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  1080
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS  1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1200
AEDEKLDKVK IAISNKEWLE YAQTSVKH                                    1228

SEQ ID NO: 58          moltype = AA  length = 1307
FEATURE                Location/Qualifiers
source                 1..1307
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 58
```

-continued

```
MTQFEGFTNL YQVSKTLRFE LIPQGKTLKH IQEQGFIEED KARNDHYKEL KPIIDRIYKT  60
YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHDYF IGRTDNLTDA  120
INKRHAEIYK GLFKAELFNG KVLKQLGTVT TTEHENALLR SFDKFTTYFS GFYRNRKNVF  180
SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV  240
FSFPFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDE TAHIIASLPH  300
RFIPLFKQIL SDRNTLSFIL EEFKSDEEVI QSFCKYKTLL RNENVLETAE ALFNELNSID  360
LTHIFISHKK LETISSALCD HWDTLRNALY ERRISELTGK ITKSAKEKVQ RSLKHEDINL  420
QEIISAAGKE LSEAFKQKTS EILSHAHAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL  480
LDWFAVDESN EVDPEFSARL TGIKLEMEPS LSFYNKARNY ATKKPYSVEK FKLNFQMPTL  540
ARGWDVNREK NNGAILFVKN GLYYLGIMPK QKGRYKALSF EPTEKTSEGF DKMYYDYFPD  600
AAKMIPKCST QLKAVTAHFQ THTTPILLSN NFIEPLEITK EIYDLNNPEK EPKKFQTAYA  660
KKTGDQKGYR EALCKWIDFT RDFLSKYTKT TSIDLSSLRP SSQYKDLGEY YAELNPLLYH  720
ISFQRIAEKE IMDAVETGKL YLFQIYNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK  780
LNGQAELFYR PKSRMKRMAA RLGEKMLNKK LKDQKTPIPD TLYQELYDYV NHRLSHDLSD  840
EARALLPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNYQ AANSPSKFNQ RVNAYLKEHP  900
ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQQFDY QKKLDNREKE RVAARQAWSV  960
VGTIKDLKQG YLSQVIHEIV DLMIHYQAVV VLENLNFGFK SKRTGIAEKA VYQQFEKMLI  1020
DKLNCLVLKD YPAEKVGGVL NPYQLTDQFT SFAKMGTQSG FLFYVPAPYT SKIDPLTGFV  1080
DPFVWKTIKN HESRKHFLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF  1140
EKNETQFDAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKG IVFRDGSNIL  1200
PKLLENDDSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCFD SRFQNPEWPM  1260
DADANGAYHI ALKGQLLLNH LKESKDLKLQ NGISNQDWLA YIQELRN                1307

SEQ ID NO: 59          moltype = AA  length = 1307
FEATURE                Location/Qualifiers
source                 1..1307
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 59
MTQFEGFTNL YQVSKTLRFE LIPQGKTLKH IQEQGFIEED KARNDHYKEL KPIIDRIYKT  60
YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHDYF IGRTDNLTDA  120
INKRHAEIYK GLFKAELFNG KVLKQLGTVT TTEHENALLR SFDKFTTYFS GFYRNRKNVF  180
SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV  240
FSFPFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKNDE TAHIIASLPH  300
RFIPLFKQIL SDRNTLSFIL EEFKSDEEVI QSFCKYKTLL RNENVLETAE ALFNELNSID  360
LTHIFISHKK LETISSALCD HWDTLRNALY ERRISELTGK ITKSAKEKVQ RSLKHEDINL  420
QEIISAAGKE LSEAFKQKTS EILSHAHAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL  480
LDWFAVDESN EVDPEFSARL TGIKLEMEPS LSFYNKARNY ATKKPYSVEK FKLNFQMPTL  540
ARGWDVNREK NNGAILFVKN GLYYLGIMPK QKGRYKALSF EPTEKTSEGF DKMYYDYFPD  600
AAKMIPKCST QLKAVTAHFQ THTTPILLSN NFIEPLEITK EIYDLNNPEK EPKKFQTAYA  660
KKTGDQKGYR EALCKWIDFT RDFLSKYTKT TSIDLSSLRP SSQYKDLGEY YAELNPLLYH  720
ISFQRIAEKE IMDAVETGKL YLFQIYNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK  780
LNGQAELFYR PKSRMKRMAH RLGEKMLNKK LKDQKTPIPD TLYQELYDYV NHRLSHDLSD  840
EARALLPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNYQ AANSPSKFNQ RVNAYLKEHP  900
ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQQFDY QKKLDNREKE RVAARQAWSV  960
VGTIKDLKQG YLSQVIHEIV DLMIHYQAVV VLENLNFGFK SKRTGIAEKA VYQQFEKMLI  1020
DKLNCLVLKD YPAEKVGGVL NPYQLTDQFT SFAKMGTQSG FLFYVPAPYT SKIDPLTGFV  1080
DPFVWKTIKN HESRKHFLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GFMPAWDIVF  1140
EKNETQFDAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKG IVFRDGSNIL  1200
PKLLENDDSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCFD SRFQNPEWPM  1260
DADANGAYHI ALKGQLLLNH LKESKDLKLQ NGISNQDWLA YIQELRN                1307

SEQ ID NO: 60          moltype = AA  length = 1107
FEATURE                Location/Qualifiers
source                 1..1107
                       mol_type = protein
                       organism = Bacillus hisashii
SEQUENCE: 60
ATRSFILKIE PNEEVKKGLW KTHEVLNHGI AYYMNILKLI RQEAIYEHHE QDPKNPKKVS  60
KAEIQAELWD FVLKMQKCNS FTHEVDKDEV FNILRELYEE LVPSSVEKKG EANQLSNKFL  120
YPLVDPNSQS GKGTASSGRK PRWYNLKIAG DPSWEEEKKK WEEDKKKDPL AKILGKLAEY  180
GLIPLFIPYT DSNEPIVKEI KWMEKSRNQS VRRLDKDMFI QALERFLSWE SWNLKVKEEY  240
EKVEKEYKTL EERIKEDIQA LKALEQYEKE RQEQLLRDTL NTNEYRLSKR GLRGWREIIQ  300
KWLKMDENEP SEKYLEVFKD YQRKHPREAG DYSVYEFLSK KENHFIWRNH PEYPYLYATF  360
CEIDKKKKDA KQQATFTLAD PINHPLWVRF EERSGSNLNK YRILTEQLHT EKLKKKLTVQ  420
LDRLIYPTES GGWEEKGKVD IVLLPSRQFY NQIFLDIEEK GKHAFTYKDE SIKFPLKGTL  480
GGARVQFDRD HLRRYPHKVE SGNVGRIYFN MTVNIEPTES PVSKSLKIHR DDFPKVVNFK  540
PKELTEWIKD SKGKKLKSGI ESLEIGLRVM SIDLGQRQAA AASIFEVVDQ KPDIEGKLFF  600
PIKGTELYAV HRASFNIKLP GETLVKSREV LRKAREDNLK LMNQKLNFLR NVLHFQQFED  660
ITEREKRVTK WISRQENSDV PLVYQDELIQ IRELMYKPYK DWVAFLKQLH KRLEVEIGKE  720
VKHWRKSLSD GRKGLYGISL KNIDEIDRTR KFLLRWSLRP TEPGEVRRLE PGQRFAIDQL  780
NHLNALKEDR LKKMANTIIM HALGYCYDVR KKKWQAKNPA CQIILFEDLS NYNPYGERSR  840
FENSRLMKWS RREIPRQVAL QGEIYGLQVG EVGAQFSSRF HAKTGSPGIR CRVVTKEKLQ  900
DNRFFKNLQR EGRLTLDKIA VLKEGDLYPD KGGEKFISLS KDRKCVTTHA DINAAQNLQK  960
RFWTRTHGFY KVYCKAYQVD GQTVYIPESK DQKQKIIEEF GEGYFILKDG VYEWVNAGKL  1020
KIKKGSSKQS SSELVDSDIL KDSFDLASEL KGEKLMLYRD PSGNVFPSDK WMAAGVFFGK  1080
```

-continued

```
LERILISKLT NQYSISTIED DSSKQSM                                  1107

SEQ ID NO: 61          moltype = DNA  length = 3987
FEATURE                Location/Qualifiers
source                 1..3987
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 61
atggccggga gcaagaagcg ccggataaag caggacacgc agttcgaggg cttcaccaac  60
ctgtaccaag tctccaagac gctccggttc gagcttatcc cgcaagggaa gaccctgaaa  120
cacatccagg aacaaggttt catcgaggag gacaaggccc gcaacgacca ctacaaggag  180
ctcaagccca taatcgatcg gatctacaag acgtacgccg accagtgcct ccaactggtg  240
cagctcgact gggagaacct gagcgccgcc attgacagct accgcaagga aaagacggag  300
gagacgcgca acgcccttat tgaggagcaa gccacctacc gcaacgccat ccacgactac  360
ttcatcgggc gcaccgacaa cctgacggac gcgatcaaca agcgccacgc ggaaatctac  420
aagggccttt tcaaggccga gctcttcaac gggaaggtcc taaaacagct cgggactgtc  480
acgacaaccg agcatgagaa cgccctcctt cgcagcttcg acaagttcac cacatacttc  540
tcgggcttct accggaaccg caagaacgtt ttcagcgccg aggacatctc caccgccatc  600
ccgcacagga tcgtccagga caacttcccc aagttcaagg agaactgcca catcttcacg  660
cgcctgatta cagccgtacc ttcacttcgt gagcacttcg agaacgtcaa aaaggccatc  720
gggatcttcg tctccacgtc catcgaggag gtattctctt tcccgttcta taaccagctc  780
ctgacccaga cgcagatcga cctctacaac cagctactgg gcggcatcag ccgggaggcc  840
gggaccgaga aaataaaggg cctcaacgaa gttctcaacc tggccatcca gaagaacgac  900
gagaccgcgc atatcatcgc atccctgccg catcgcttca ttcctttgtt caagcagata  960
ttgagcgacc ggaacaccct ctcgttcatc ctcgaagaat tcaagagcga cgaggaggtc  1020
attcagtctt tctgcaagta caagacgctc ctacggaatg agaatgtgct ggagaccgcg  1080
gaggcactct tcaatgagct gaactccatt gacctgaccc acatcttcat tagccacaag  1140
aaactggaga cgatctccag cgccctgtgc gaccactggg acactctccg caacgccctc  1200
tacgaacgcc ggatctccga acttaccggc aagataacta agtcggctaa ggagaaggtg  1260
caacggagcc tcaagcacga ggacatcaac cttcaggaaa tcatctcagc cgcgggcaag  1320
gagctgagcg aggcgtttaa gcagaaaaca tcggagatac tgagccacgc gcacgcggcc  1380
ctggatcaac cgctgccgac gactctcaag aagcaagagg agaaggaaat ccttaagtcc  1440
cagctcgact cgctgctcgg cctctatcac ttgctcgact ggttcgcggt tgatgagtcc  1500
aacgaggtgg acccggagtt ctccgcgcgc ctcacgggta ttaagctgga gatggagcca  1560
agcttaagct tctacaacaa ggcccgcaac tacgcgacca aaaaaccgta ctcagtcgag  1620
aaattcaagc tgaatttcca gatgcctaca ttggcgaggg ggtgggacgt gaaccgcgag  1680
aagaacaatg gagccatcct gttcgtcaaa aatgggttgt actacctggg catcatgccc  1740
aagcagaagg gccgttacaa ggccctgtca ttcgagccta ccgagaagac ctcggagggc  1800
ttcgacaaga tgtactacga ctatttcccg gacgccgcca agatgatccc gaagtgctcc  1860
acgcagctca aagccgtcac ggcccacttc cagacgcata ccacgccgat acttctgagc  1920
aacaacttca ttgagccgct agagatcacg aaggagatat acgacctaaa caaccccgaa  1980
aaggagccca agaagttcca gacagcctac gctaagaaga caggtgatca gaagggatat  2040
agggaggcac tctgcaagtg gatcgacttc acgcgcgact tcctgtcgaa atatacaaag  2100
acgaccagca ttgacctaag ttctctccgc ccatcctccc agtacaagga tctgggcgag  2160
tattatgcgg agctgaaccc attgctgtac cacatcagct tccagaggat cgccgagaag  2220
gagattatgg acgcggtgga gacggggaaa ctatacctgt tccaaatata taacaaggac  2280
ttcgctaaag ggcaccacgg gaagcccaac ctgcacacac tctactggac gggcttgttt  2340
tcgccagaaa atttggccaa gacttcgatc aagctcaacg gccaggcgga gttgttttac  2400
cgtcccaagt ctcgcatgaa gcgcatggcg catcgcctcg gagagaaaat gcttaacaag  2460
aagctcaagg atcagaagac gcccatacct gatacgttgt accaggaatt gtacgactac  2520
gtgaaccacc gcctatcgca cgacctctca gacgaggccc gcgccctcct cccaaacgtg  2580
attactaagg aggtttccca tgaaataatc aaggaccgac ggttcaccag cgacaaattt  2640
tttttccacg tgcctatcac gctcaattac caggcggcca actccccatc gaagttcaac  2700
cagcgcgtga acgcctacct taaggagcac ccggagaccc caatcatcgg gatcgaccgt  2760
ggcgagcgga acctgatcta tattacggtg atcgatagca ccgggaagat cctggagcag  2820
cgctccctga acacaatcca gcagtttgac taccagaaga aactcgacaa ccgggagaag  2880
gagcgcgtcg cagcccggca agcatggagt gtggtcggca ccataaagga cctgaaacag  2940
ggttacctaa gtcaagttat ccacgagatc gttgacctga tgatacacta tcaagccgta  3000
gtcgtgctgg agaacctcaa cttcgggttt aagtccaagc gcaccggcat cgcggagaag  3060
gcggtgtacc agcagttcga gaagatgctg atcgacaagc tgaactgcct ggtgctcaag  3120
gactaccctg cggagaaggt cggcggggtc ttgaacccgt accagctaac cgaccagttc  3180
acgagcttcg ccaaaatggg cacgcagtcc ggattcttgt tttatgtccc ggctccatat  3240
acaagtaaga tcgacccgct gacagggttt gttgacccat tcgtgtggaa gaccatcaag  3300
aaccacgaga gcaggaaaca cttcttagag ggcttcgact tcctgcatta cgacgttaag  3360
acaggcgact tcatcctgca cttcaagatg aaccgcaacc tgtcgttcca gaggggcctg  3420
cccggcttca tgcccgcctg ggatatcgtc tttgagaaga atgagacgca gttcgacgcg  3480
aaggggacgc cgttcatcgc tggaaagcgg atcgtgccgg tcatcgagaa ccaccgcttc  3540
acgggtcgct accgagattt ataccccgcc aacgaactaa ttgcgctgct ggaggagaag  3600
gggatcgtgt tccgagatgg cagcaacatt ctcccgaagc tgctggagaa cgacgactcg  3660
cacgctattg acacgatggt cgccctcata cggagcgtgc ttcagatgcg gaacagtaac  3720
gctgccacgg gcgaggacta cattaactcc cccgtccgcg acctcaacgg ggtctgcttc  3780
gatagccgct tccagaaccc ggagtggcct atggatgcgg acgcgaacgg ggcctaccac  3840
atcgccctca agggccaact cctgctcaac cacttgaagg aaagcaaaga cctcaaattg  3900
cagaatggca tcagtaacca ggactggctc gcgtacatcc aggaactgag aaacgggtcc  3960
aagaagcggc gtatcaagca agattga                                  3987
```

-continued

```
SEQ ID NO: 62          moltype = DNA  length = 3987
FEATURE                Location/Qualifiers
source                 1..3987
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 62
atggcgggaa gcaaaaagcg ccggattaag caagacacgc agttcgaggg cttcacgaac  60
ctctaccaag tcagcaagac cctccggttc gagctgatac cacagggaaa gacgctcaag  120
cacatccagg aacagggctt catcgaggag gacaaggcgc gcaacgacca ctacaaggag  180
ttgaaaccga tcatcgaccg catctacaag acgtacgccg accagtgcct ccagctcgtg  240
cagctcgact gggagaacct ctccgccgcc attgactcgt accggaagga gaagactgag  300
gagacccgca acgccctgat cgaggagcaa gcaacctacc ggaacgccat ccacgactac  360
ttcatcggcc gcaccgacaa cctcaccgac gcgatcaaca agcggcacgc ggagatatac  420
aaagggctgt tcaaggcgga gctgttcaac ggcaaggtgc tcaagcagct agggacggtg  480
accacgaccg agcacgagaa cgcgctcctc cgcagcttcg acaagttcac cacctacttc  540
agcggcttct accggaaccg caagaatgtg ttcagcgcgg aggacatcag cacggccatc  600
ccgcaccgca tcgtccagga caacttcccg aagttcaagg agaactgcca catcttcacc  660
cgcctgataa ccgccgtccc ctccctgcgg gagcacttcg agaacgtcaa aaaggcaatt  720
gggatcttcg tctcgaccag cattgaggag gtgttcagct tcccсttcta caaccagctc  780
ctcacccaga cgcagatcga cctgtacaat cagttgctcg gcgggataag ccgcgaggcg  840
ggaaccgaaa aaatcaaggg gctgaacgaa gtgttgaacc tcgccatcca gaagaacgac  900
gagaccgcgc acatcatcgc ctccctgccc caccggttca tcccgctgtt caagcagatc  960
ctctctgacc ggaacaccct gtccttcatt cttgaggagt tcaagtcgga cgaggaggtc  1020
atccagagct tctgcaagta caagacgctg ctacggaacg agaacgtgct ggagacggcg  1080
gaggcactgt tcaacgagct aaacagcatc gacctcacgc acatcttcat cagtcacaag  1140
aaactggaga ccatctcctc cgcgctgtgc gaccactggg acacgctcag gaacgcgctc  1200
tacgagcgcc gaatcagtga gctgacgggc aagatcacga agtccgcgaa ggagaaggtg  1260
cagcggtccc tcaagcacga ggacatcaac ctccaggaga tcatctcagc ggctgggaaa  1320
gagctgtccg aggcgttcaa gcagaaaacg agcgaaatcc tgtcccacgc gcacgcggcc  1380
ctggatcagc ctctgccgac gaccctcaag aaacaagaag aaaaggaaat cctcaagtcg  1440
cagctcgact cgctgctggg cctgtaccat ctcctcgact ggttcgccgt ggacgagagc  1500
aacgaggtgg accccgagtt ctccgcgcgg cttacgggga tcaagctgga gatggagccc  1560
agcctgtcct tctacaacaa ggcgcgcaac tacgccacca agaagcccta cagcgtggag  1620
aagttcaagc tcaacttcca gatgcccact ctcgcacgtg ggtgggacgt caaccgcgaa  1680
aaaaataatg gggcgatcct gttcgtcaag aacggcctgt actacttggg catcatgccg  1740
aaacagaagg gccgctacaa ggccctgagc ttcgaaccga ccgagaaaac gagcgagggg  1800
ttcgacaaga tgtactacga ctacttcccc gacgccgcga agatgattcc aaagtgctcc  1860
acgcagctta aggccgtgac ggcccacttc cagacgcaca cgaccccgat cctcctcagc  1920
aacaacttca tcgagcccct ggagatcacg aaggagatat acgacctgaa caacccggag  1980
aaggagccca agaaattcca gaccgcctac gccaagaaga caggcgacca aaagggttac  2040
agggaggccc tctgcaagtg gatcgacttc actagggact tcctgtccaa gtacaccaag  2100
actacctcta tcgacctgtc cagcctccgc ccgtcgtccc agtacaaaga tttgggcgag  2160
tattacgcgg agctgaaccc actgctctac cacatcagct tccagcgcat cgcggagaag  2220
gagatcatgg acgcagtgga gacgggcaag ctatacctat ttcagatata caacaaagac  2280
ttcgctaagg gacaccacgg caagcctaac ctgcacaccc tctactggac ggggctcttc  2340
agcccggaga acctcgccaa gacctcgatc aagctcaacg gccaggccga gctgttctac  2400
cggcccaagt cccgcatgaa gcggatggcc caccggctcg ggagaaaat gctcaacaag  2460
aaattgaagg accaaaaaac gccgataccc gacaccctat accaggagct gtacgactat  2520
gtgaaccacc gcctgagcca cgacctcagc gacgaggcgc gggccctcct gccgaacgtc  2580
atcacaaagg aggtcagcca cgagatcatc aaggaccggc gcttcacctc cgacaagttt  2640
ttctttcacg tgcccatcac gctcaactac caggccgcca actcgccgtc caagttcaac  2700
cagcgcgtga acgcctacct caaggagcac cccgagaccc cgatcatcgg gattgaccga  2760
ggggagcgga acctcatcta catcaccgtc atcgacagca ccgggaagat ccttgaacag  2820
cggtcgctca acaccatcca gcagttcgac taccagaaga aactcgacaa ccgggagaag  2880
gagagagtgg cggcccgcca ggcttggtcc gtcgtcggga cgattaagga cttgaaacaa  2940
ggttacctgt cgcaagtgat ccacgagatc gttgacctga tgatccacta ccaagccgtc  3000
gtggtcctgg agaacctcaa cttcggcttc aagagcaaac gaaccggcat cgcggagaag  3060
gccgtgtacc agcagttcga aaaaatgctg atcgacaagc tgaactgcct cgtgctcaag  3120
gactaccccg ctgagaaggt cggcggggtg ctgaacccgt accagctcac tgaccagttc  3180
accagcttcg caaagatggg cacccagtcc ggcttcctgt tctacgtgcc tgcgccatac  3240
acctcgaaga tcgacccgct caccgggttc gtggacccct tcgtctggaa gaccatcaag  3300
aaccacgaga gccgcaagca cttcctggag ggcttcgact tcctccacta cgacgtcaag  3360
accggggact tcatcctgca cttcaagatg aaccgcaacc tcagtttcca gcgcggcctg  3420
ccggggttca tgcccgcttg ggatatagtc ttcgagaaga atgagacgca gttcgacgcg  3480
aagggcaccc cgttcatcgc cgggaagcgc atcgtgccgg tcatcgagaa ccaccggttc  3540
accgggcgct accgcgacct atacccggcg aacgagttga tcgccctcct ggaggagaag  3600
ggcatcgtgt tccgcgacgg ctccaacatc ctcccgaagc tgctcgaaaa cgacgactcc  3660
cacgccatcg acacgatggt cgcgctgatc cggtcggtgc tccagatgcg gaactccaac  3720
gccgcgacgg gcgaggacta catcaacagt ccggtccgcg atctgaacgg cgtctgcttc  3780
gactcccggt tccagaaccc cgagtggccg atggacgcgg acgcgaacgg cgcataccac  3840
atcgccctaa aagggcaatt gctgctcaac cacctcaagg aatccaaaga cctaaagctc  3900
cagaacggca tctccaacca ggactggctg gcgtacatcc aggaactgcg gaacgggagc  3960
aaaaaacgtc ggatcaagca agattga                                      3987

SEQ ID NO: 63          moltype = DNA  length = 3987
FEATURE                Location/Qualifiers
```

-continued

```
source                 1..3987
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 63
atggcgggct ccaagaaacg ccggattaag caagataccc agttcgaggg gttcacgaac  60
ctctaccaag tgagcaagac cctccgattc gaactgattc ctcaggggaa gaccctcaag  120
cacatccagg agcaagggtt catcgaggag gacaaggcgc ggaacgacca ctacaaggaa  180
ctcaaaccca tcatcgaccg catctacaag acctacgccg atcagtgcct ccagctcgtg  240
cagttggact gggagaacct cagcgcggcc attgactcct accggaagga gaaaacggag  300
gagacgcgca acgcgctcat cgaggaacag gcaacctatc gcaacgccat ccacgactac  360
ttcatcggga ggactgacaa cctcactgac gcgattaaca agcgccacgc ggagatatac  420
aagggactct tcaaagcgga gctgtttaac ggcaaggttc tcaagcaact cggcactgtg  480
accacgaccg agcatgagaa cgccctgctc cgctccttcg acaagttcac cacctacttc  540
tccgggttct accgcaaccg caagaatgtc ttcagcgcgg aggacatcag cacggccatt  600
ccacatcgaa tcgtccaaga taacttcccg aagttcaagg agaactgcca catcttcacc  660
cgactcatta ctgctgtacc gtcgttacgc gaacacttcg agaacgtcaa gaaggcaatt  720
ggaatcttcg tctctacgtc aatagaggag gtgttcagct tccctttcta caaccagctc  780
cttacgcaga cccagataga cctgtacaat cagctcctcg gtgggatcag ccgggaggcg  840
gggactgaga agattaaagg gctcaacgag gtcttgaacc tggccatcca aaaaaacgat  900
gagacggcgc acatcatcgc ctcgctgccc caccggttca tcccgctgtt caagcagatc  960
ctcagtgaca ggaacacctt gagctttatc ctagaggagt tcaagagcga cgaggaggtg  1020
atccagagct tctgcaagta caaaaccctg ctgaggaacg agaacgtcct ggagacggcg  1080
gaggcgctgt tcaacgagct gaactctatc gacttaactc acatattcat ctcgcacaag  1140
aagctggaga ctattagctc tgcactctgc gaccactggg acaccctccg caacgcgctc  1200
tacgagcgcc gcatctcgga gctgaccggg aagatcacca aatccgcgaa ggaaaaggtc  1260
cagcgttccc tcaaacacga ggatattaac ttacaggaga ttatctcagc ggctgggaag  1320
gagttgtcag aggcgttcaa gcagaaaact tccgagatcc tgagccacgc gcacgcagcg  1380
ctcgaccagc ctctgcccac caccctcaaa aagcaggaag aaaaagagat cctcaagagc  1440
cagttggact ccctgctggg gctctatcac cttctcgact ggttcgccgt cgatgagtcg  1500
aacgaggtgg accccgagtt ctccgcccgg ctgaccggca tcaagctaga gatggagccg  1560
tccctcagct tctacaataa ggcccgcaac tacgcgacca aaaaacccta cagcgtggag  1620
aagttcaagc tgaacttcca gatgccgacc ttagcacgcg gttgggacgt aaacagggag  1680
aagaacaatg gagccatcct gttcgtcaag aacgggcttt actacctcgg gataatgccc  1740
aagcagaagg gccgctacaa ggccctttcc ttcgagccga cggagaaaac ctccgagggg  1800
ttcgacaaga tgtactacga ctacttcccc gacgccgcca agatgatccc gaagtgctca  1860
acgcagctaa aagccgtgac cgcccacttc cagacccaca cgacgccgat cctgctgagc  1920
aacaacttca tcgagcccct tgagatcact aaggagatat acgacctgaa caaccccgag  1980
aaggagccca agaagtttca aaccgcctac gccaaaaaaa ctggcgacca aaagggctac  2040
agggaggcgc tgtgtaagtg gatcgacttc acacgcgact tcctttcgaa gtatacgaag  2100
acaacctcta ttgacctgag cagcctgcgt cctagctccc agtacaaaga tttgggcgag  2160
tactacgcgg agcttaatcc actactctac cacatctcat tccagcgcat cgctgagaag  2220
gaaatcatgg acgcggtgga gacaggcaaa ctgtacctct tccagatata caacaaagac  2280
ttcgctaagg ggcaccacgg gaagcccaac cttcatacgc tctactggac gggcctattc  2340
agccccgaaa atctggccaa gacctccatc aagctgaacg gccaagcgga gctgttctac  2400
agacccaaga gccggatgaa gcggatggcc cacaggctcg gcgagaaaat gcttaacaaa  2460
aagttgaagg accagaaaac ccctatcccc gacaccctct accaggaact gtacgactac  2520
gtgaaccaca ggctctcgca cgacctttcc gacgaggccc gtgccctact cccgaacgtc  2580
attaccaaag aggtttcgca cgagatcatc aaggaccggc ggttcacgag cgacaagttt  2640
ttctttcacg tccccatcac ccttaactac caggcggcca actccccatc caagttcaac  2700
cagcgtgtga atgcctacct caaggagcac ccagagaccc cgatcattgg gatcgaccgg  2760
ggcgagcgga acctgatcta catcaccgtc atcgactcga cgggcaagat tcttgagcag  2820
agatcgttga ataccataca gcagttcgac taccagaaga aactcgacaa ccgcgagaag  2880
gagcgcgtgg cggcccgcca ggcgtggtcc gtcgttggga cgattaagga cttgaaacaa  2940
ggttatctgt cccaagtcat ccacgagatc gttgatctga tgatccacta tcaggcagtg  3000
gtggtgctgg agaatctcaa cttcggcttc aagagtaagc ggacgggaat cgccgagaag  3060
gccgtgtacc agcagttcga gaagatgctg atcgacaagc tcaactgcct tgtgctgaaa  3120
gactacccgg ccgagaaggt cggcggcgtc ctcaacccgt accaacttac cgaccagttc  3180
acctccttcg ccaagatggg cactcagtcc gggttcttgt tctacgtccc cgcaccttac  3240
acctctaaga tcgaccctct gactggcttc gtagatccat tcgtgtggaa gaccattaag  3300
aaccacgaga gccgcaagca cttcctggag ggcttcgact tcctgcacta cgacgtgaag  3360
accggggact tcatccttca cttcaagatg aaccggaacc tcagcttcca gcggggcctg  3420
ccggggttca tgcccgcctg ggacatcgtg ttcgagaaga acgagaccca gttcgacgcg  3480
aagggcacgc ccttcatcgc cgggaagcgt atcgtgccgg tgatcgagaa ccatcgtttc  3540
acgggtcgct accgtgacct ctacccggcg aacgagctta tcgcactcct ggaggagaag  3600
ggcatcgtct tccgggacgg ctccaacatc ctcccgaaac tgctggaaaa cgacgactct  3660
cacgccatcg acacgatggt ggccctcatc cggtccgtgc tccaaatgcg gaacagcaac  3720
gccgccaccg gtgaggacta catcaacagc ccggtccggg atctgaacgg ggtgtgcttc  3780
gattcgcggt tccagaatcc tgagtggccg atggacgcgg atgcaaacgg ggcgtaccac  3840
atcgcgctca agggccagtt acttctgaac caccttaagg agtctaaaga tttgaaactc  3900
cagaacggga tctcgaacca ggactggctg gcctacatcc aagagttgcg gaacggcagc  3960
aagaagcggc ggattaagca agattag                                     3987

SEQ ID NO: 64          moltype = AA  length = 228
FEATURE                Location/Qualifiers
source                 1..228
                       mol_type = protein
```

-continued

```
                       organism = Rattus norvegicus
SEQUENCE: 64
SSETGPVAVD PTLRRRIEPH EFEVFFDPRE LRKETCLLYE INWGGRHSIW RHTSQNTNKH  60
VEVNFIEKFT TERYFCPNTR CSITWFLSWS PCGECSRAIT EFLSRYPHVT LFIYIARLYH  120
HADPRNRQGL RDLISSGVTI QIMTEQESGY CWRNFVNYSP SNEAHWPRYP HLWVRLYVLE  180
LYCIILGLPP CLNILRRKQP QLTFFTIALQ SCHYQRLPPH ILWATGLK               228

SEQ ID NO: 65          moltype = AA  length = 199
FEATURE                Location/Qualifiers
source                 1..199
                       mol_type = protein
                       organism = Homo sapiens
SEQUENCE: 65
MEASPASGPR HLMDPHIFTS NFNNGIGRHK TYLCYEVERL DNGTSVKMDQ HRGFLHNQAK  60
NLLCGFYGRH AELRFLDLVP SLQLDPAQIY RVTWFISWSP CFSWGCAGEV RAFLQENTHV  120
RLRIFAARIY DYDPLYKEAL QMLRDAGAQV SIMTYDEFKH CWDTFVDHQG CPFQPWDGLD  180
EHSQALSGRL RAILQNQGN                                               199

SEQ ID NO: 66          moltype = DNA  length = 621
FEATURE                Location/Qualifiers
source                 1..621
                       mol_type = other DNA
                       organism = Petromyzon marinus
SEQUENCE: 66
acagatgcag agtatgtgag aattcacgaa aagctggaca tctatacctt caagaagcag  60
ttctttaaca ataagaagtc tgtgagccat aggtgctacg tgctgttcga gctgaagaga  120
aggggtgaaa gaagggcatg tttttggggg tatgctgtga acaagcccca gtctggaact  180
gagagaggca ttcacgccga aattttcagc atcagaaagg tggaggaata cctgagggat  240
aaccctggac agtttacaat taattggtat tctagctggt ctccatgcgc tgactgtgcc  300
gagaagatcc tggaatggta caaccaggag ctgagaggaa atggccatac cctgaagatt  360
tgggcctgca agctgtacta tgaaaagaac gcaagaaatc agatcggact gtggaacctg  420
agggataatg gtgtggggct gaacgtgatg gtgtccgagc actatcagtg ctgtagaaag  480
attttcattc agtcctcaca taatcagctg aacgagaata gatggctgga aaagactctg  540
aagagggctg agaagagaag gtccgaactg tcaattatga tccaggtgaa gatcctgcac  600
accactaagt cacctgccgt g                                            621

SEQ ID NO: 67          moltype = AA  length = 160
FEATURE                Location/Qualifiers
source                 1..160
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 67
FERNYDPREL RKETYLLYEI KWGKSGKLWR HWCQNNRTQH AEVYFLENIF NARRFNPSTH  60
CSITWYLSWS PCAECSQKIV DFLKEHPNVL EIYVARLYYH EDERNRQGLR DLVNSGVTIR  120
IMDLPDYNYC WKTFVSDQGG DEDYWPGHFA PWIKQYSLKL                        160

SEQ ID NO: 68          moltype = AA  length = 229
FEATURE                Location/Qualifiers
source                 1..229
                       mol_type = protein
                       organism = Rattus norvegicus
SEQUENCE: 68
MSSETGPVAV DPTLRRRIEP HEFEVFFDPR ELRKETCLLY EINWGGRHSI WRHTSQNTNK  60
HVEVNFIEKF TTERYFCPNT RCSITWFLSW SPCGECSRAI TEFLSRYPHV TLFIYIARLY  120
HHADPRNRQG LRDLISSGVT IQIMTEQESG YCWRNFVNYS PSNEAHWPRY PHLWVRLYVL  180
ELYCIILGLP PCLNILRRKQ PQLTFFTIAL QSCHYQRLPP HILWATGLK              229

SEQ ID NO: 69          moltype = AA  length = 198
FEATURE                Location/Qualifiers
source                 1..198
                       mol_type = protein
                       organism = Homo sapiens
SEQUENCE: 69
MDSLLMNRRK FLYQFKNVRW AKGRRETYLC YVVKRRDSAT SFSLDFGYLR NKNGCHVELL  60
FLRYISDWDL DPGRCYRVTW FTSWSPCYDC ARHVADFLRG NPNLSLRIFT ARLYFCEDRK  120
AEPEGLRRLH RAGVQIAIMT FKDYFYCWNT FVENHERTFK AWEGLHENSV RLSRQLRRIL  180
LPLYEVDDLR DAFRTLGL                                                198

SEQ ID NO: 70          moltype = AA  length = 197
FEATURE                Location/Qualifiers
REGION                 1..197
                       note = Synthetic polypeptide
source                 1..197
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 70
```

-continued

```
MDSLLMNRRE FLYQFKNVRW AKGRRETYLC YVVKRRDSAT SFSLDFGYLR NKNGCHVELL  60
FLRYISDWDL DPGRCYRVTW FISWSPCYDC ARHVADFLRG NPNLSLRIFT ARLYFCEDRK  120
AEPEGLRRLH RAGVQIAIMT FKDYFYCWNT FVENHGRTFK AWEGLHENSV RLSRQLRRIL  180
LPLYEVDDLR DAFRTCT                                                 197

SEQ ID NO: 71          moltype = AA  length = 207
FEATURE                Location/Qualifiers
source                 1..207
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 71
TDAEYVRIHE KLDIYTFKKQ FSNNKKSVSH RCYVLFELKR RGERRACFWG YAVNKPQSGT  60
ERGIHAEIFS IRKVEEYLRD NPGQFTINWY SSWSPCADCA EKILEWYNQE LRGNGHTLKI  120
WVCKLYYEKN ARNQIGLWNL RDNGVGLNVM VSEHYQCCRK IFIQSSHNQL NENRWLEKTL  180
KRAEKRRSEL SIMFQVKILH TTKSPAV                                      207

SEQ ID NO: 72          moltype = AA  length = 228
FEATURE                Location/Qualifiers
source                 1..228
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 72
SSKTGPVAVD PTLRRRIEPH EFEVFFDPRE LRKETCLLYE INWGGRHSIW RHTSQNTNKH  60
VEVNFIEKFT TERYFCPNTR CSITWFLSWS PCGECSRAIT EFLSRYPNVT LFIYIARLYH  120
LANPRNRQGL RDLISSGVTI QIMTEQESGY CWHNFVNYSP SNESHWPRYP HLWVRLYVLE  180
LYCIILGLPP CLNILRRKQS QLTSFTIALQ SCHYQRLPPH ILWATGLK               228

SEQ ID NO: 73          moltype = AA  length = 162
FEATURE                Location/Qualifiers
source                 1..162
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 73
SFERNYDPRE LRKETYLLYE IKWGKSGKLW RHWCQNNRTQ HAEVYFLENI FNARRFNPST  60
HCSITWYLSW SPCAECSQKI VDFLKEHPNV NLEIYVARLY YPENERNRQG LRDLVNSGVT  120
IRIMDLPDYN YCWKTFVSDQ GGDEDYWPGH FAPWIKQYSL KL                     162

SEQ ID NO: 74          moltype = AA  length = 166
FEATURE                Location/Qualifiers
source                 1..166
                       mol_type = protein
                       organism = Escherichia coli
SEQUENCE: 74
SEVEFSHEYW MRHALTLAKR AWDEREVPVG AVLVHNNRVI GEGWNRPIGR HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTLEPCVMCA GAMIHSRIGR VVFGARDAKT GAAGSLMDVL  120
HHPGMNHRVE ITEGILADEC AALLSDFFRM RRQEIKAQKK AQSSTD                 166

SEQ ID NO: 75          moltype = AA  length = 166
FEATURE                Location/Qualifiers
source                 1..166
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 75
SEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNNRVI GEGWNRAIGL HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL  120
HYPGMNHRVE ITEGILADEC AALLCYFFRM PRQVFNAQKK AQSSTD                 166

SEQ ID NO: 76          moltype = AA  length = 166
FEATURE                Location/Qualifiers
source                 1..166
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 76
SEVEFSHEYW MRHALTLAKR AWDEREVPVG AVLVLNNRVI GEGWNRSIGL HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL  120
HYPGMNHRVE ITEGILADEC AALLCYFFRM RRQVFNAQKK AQSSTD                 166

SEQ ID NO: 77          moltype = AA  length = 166
FEATURE                Location/Qualifiers
source                 1..166
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 77
SEVEFSHEYW MRHALTLAKR ALDEREVPVG AVLVLNNRVI GEGWNRAIGL HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL  120
HYPGMNHRVE ITEGILADEC NALLCYFFRM RRQVFNAQKK AQSSTD                 166
```

-continued

```
SEQ ID NO: 78          moltype = AA  length = 166
FEATURE                Location/Qualifiers
source                 1..166
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 78
SEVEFSHEYW MRHALTLAKR ALDEREVPVG AVLVLNNRVI GEGWNRAIGL HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL  120
HYPGMNHRVE ITEGILADEC NALLCYFFRM PRQVFNAQKK AQSSTD                 166

SEQ ID NO: 79          moltype = AA  length = 1763
FEATURE                Location/Qualifiers
source                 1..1763
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 79
SEVEFSHEYW MRHALTLAKR AWDEREVPVG AVLVHNNRVI GEGWNRPIGR HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTLEPCVMCA GAMIHSRIGR VVFGARDAKT GAAGSLMDVL  120
HHPGMNHRVE ITEGILADEC AALLSDFFRM RRQEIKAQKK AQSSTDSGGS SGGSSGSETP  180
GTSESATPES SGGSSGGSSE VEFSHEYWMR HALTLAKRAR DEREVPVGAV LVLNNRVIGE  240
GWNRAIGLHD PTAHAEIMAL RQGGLVMQNY RLIDATLYVT FEPCVMCAGA MIHSRIGRVV  300
FGVRNAKTGA AGSLMDVLHY PGMNHRVEIT EGILADECAA LLCYFFRMPR QVFNAQKKAQ  360
SSTDSGGSSG GSSGSETPGT SESATPESSG GSSGGSDKKY SIGLAIGTNS VGWAVITDEY  420
KVPSKKFKVL GNTDRHSIKK NLIGALLFDS GETAEATRLK RTARRRYTRR KNRICYLQEI  480
FSNEMAKVDD SFFHRLEESF LVEEDKKHER HPIFGNIVDE VAYHEKYPTI YHLRKKLVDS  540
TDKADLRLIY LALAHMIKFR GHFLIEGDLN PDNSDVDKLF IQLVQTYNQL FEENPINASG  600
VDAKAILSAR LSKSRRLENL IAQLPGEKKN GLFGNLIALS LGLTPNFKSN FDLAEDAKLQ  660
LSKDTYDDDL DNLLAQIGDQ YADLFLAAKN LSDAILLSDI LRVNTEITKA PLSASMIKRY  720
DEHHQDLTLL KALVRQQLPE KYKEIFFDQS KNGYAGYIDG GASQEEFYKF IKPILEKMDG  780
TEELLVKLNR EDLLRKQRTF DNGSIPHQIH LGELHAILRR QEDFYPFLKD NREKIEKILT  840
FRIPYYVGPL ARGNSRFAWM TRKSEETITP WNFEEVVDKG ASAQSFIERM TNFDKNLPNE  900
KVLPKHSLLY EYFTVYNELT KVKYVTEGMR KPAFLSGEQK KAIVDLLFKT NRKVTVKQLK  960
EDYFKKIECF DSVEISGVED RFNASLGTYH DLLKIIKDKD FLDNEENEDI LEDIVLTLTL  1020
FEDREMIEER LKTYAHLFDD KVMKQLKRRR YTGWGRLSRK LINGIRDKQS GKTILDFLKS  1080
DGFANRNFMQ LIHDDSLTFK EDIQKAQVSG QGDSLHEHIA NLAGSPAIKK GILQTVKVVD  1140
ELVKVMGRHK PENIVIEMAR ENQTTQKGQK NSRERMKRIE EGIKELGSQI LKEHPVENTQ  1200
LQNEKLYLYY LQNGRDMYVD QELDINRLSD YDVDHIVPQS FLKDDSIDNK VLTRSDKNRG  1260
KSDNVPSEEV VKKMKNYWRQ LLNAKLITQR KFDNLTKAER GGLSELDKAG FIKRQLVETR  1320
QITKHVAQIL DSRMNTKYDE NDKLIREVKV ITLKSKLVSD FRKDFQFYKV REINNYHHAH  1380
DAYLNAVVGT ALIKKYPKLE SEFVYGDYKV YDVRKMIAKS EQEIGKATAK YFFYSNIMNF  1440
FKTEITLANG EIRKRPLIET NGETGEIVWD KGRDFATVRK VLSMPQVNIV KKTEVQTGGF  1500
SKESILPKRN SDKLIARKKD WDPKKYGGFD SPTVAYSVLV VAKVEKGKSK KLKSVKELLG  1560
ITIMERSSFE KNPIDFLEAK GYKEVKKDLI IKLPKYSLFE LENGRKRMLA SAGELQKGNE  1620
LALPSKYVNF LYLASHYEKL KGSPEDNEQK QLFVEQHKHY LDEIIEQISE FSKRVILADA  1680
NLDKVLSAYN KHRDKPIREQ AENIIHLFTL TNLGAPAAFK YFDTTIDRKR YTSTKEVLDA  1740
TLIHQSITGL YETRIDLSQL GGD                                          1763

SEQ ID NO: 80          moltype = AA  length = 1565
FEATURE                Location/Qualifiers
source                 1..1565
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 80
SEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNNRVI GEGWNRAIGL HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTFEPCVMCA GAMIHSRIGR VVFGVRNSKR GAAGSLMNVL  120
NYPGMNHRVE ITEGILADEC AALLCDFYRM PRQVFNAQKK AQSSINSGGS SGGSSGSETP  180
GTSESATPES SGGSSGGSDK KYSIGLAIGT NSVGWAVITD EYKVPSKKFK VLGNTDRHSI  240
KKNLIGALLF DSGETAEATR LKRTARRRYT RRKNRICYLQ EIFSNEMAKV DDSFFHRLEE  300
SFLVEEDKKH ERHPIFGNIV DEVAYHEKYP TIYHLRKKLV DSTDKADLRL IYLALAHMIK  360
FRGHFLIEGD LNPDNSDVDK LFIQLVQTYN QLFEENPINA SGVDAKAILS ARLSKSRRLE  420
NLIAQLPGEK KNGLFGNLIA LSLGLTPNFK SNFDLAEDAK LQLSKDTYDD DLDNLLAQIG  480
DQYADLFLAA KNLSDAILLS DILRVNTEIT KAPLSASMIK RYDEHHQDLT LLKALVRQQL  540
PEKYKEIFFD QSKNGYAGYI DGGASQEEFY KFIKPILEKM DGTEELLVKL NREDLLRKQR  600
TFDNGSIPHQ IHLGELHAIL RRQEDFYPFL KDNREKIEKI LTFRIPYYVG PLARGNSRFA  660
WMTRKSEETI TPWNFEEVVD KGASAQSFIE RMTNFDKNLP NEKVLPKHSL LYEYFTVYNE  720
LTKVKYVTEG MRKPAFLSGE QKKAIVDLLF KTNRKVTVKQ LKEDYFKKIE CFDSVEISGV  780
EDRFNASLGT YHDLLKIIKD KDFLDNEENE DILEDIVLTL TLFEDREMIE ERLKTYAHLF  840
DDKVMKQLKR RRYTGWGRLS RKLINGIRDK QSGKTILDFL KSDGFANRNF MQLIHDDSLT  900
FKEDIQKAQV SGQGDSLHEH IANLAGSPAI KKGILQTVKV VDELVKVMGR HKPENIVIEM  960
ARENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY  1020
VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW  1080
RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGFIKRQLVE TRQITKHVAQ ILDSRMNTKY  1140
DENDKLIREV KVITLKSKLV SDFRKDFQFY KVREINNYHH AHDAYLNAVV GTALIKKYPK  1200
LESEFVYGDY KVYDVRKMIA KSEQEIGKAT AKYFFYSNIM NFFKTEITLA NGEIRKRPLI  1260
ETNGETGEIV WDKGRDFATV RKVLSMPQVN IVKKTEVQTG GFSKESILPK RNSDKLIARK  1320
```

-continued

```
KDWDPKKYGG FDSPTVAYSV LVVAKVEKGK SKKLKSVKEL LGITIMERSS FEKNPIDFLE  1380
AKGYKEVKKD LIIKLPKYSL FELENGRKRM LASAGELQKG NELALPSKYV NFLYLASHYE  1440
KLKGSPEDNE QKQLFVEQHK HYLDEIIEQI SEFSKRVILA DANLDKVLSA YNKHRDKPIR  1500
EQAENIIHLF TLTNLGAPAA FKYFDTTIDR KRYTSTKEVL DATLIHQSIT GLYETRIDLS  1560
QLGGD                                                              1565

SEQ ID NO: 81          moltype = AA  length = 1565
FEATURE                Location/Qualifiers
source                 1..1565
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 81
SEVEFSHEYW MRHALTLAKR ARDEREVPVG AVLVLNNRVI GEGWNRAIGL HDPTAHAEIM  60
ALRQGGLVMQ NYRLYDATLY STFEPCVMCA GAMIHSRIGR VVFGVRNAKT GAAGSLMDVL  120
HHPGMNHRVE ITEGILADEC AALLCRFFRM PRRVFNAQKK AQSSTDSGGS SGGSSGSETP  180
GTSESATPES SGGSSGGSDK KYSIGLAIGT NSVGWAVITD EYKVPSKKFK VLGNTDRHSI  240
KKNLIGALLF DSGETAEATR LKRTARRRYT RRKNRICYLQ EIFSNEMAKV DDSFFHRLEE  300
SFLVEEDKKH ERHPIFGNIV DEVAYHEKYP TIYHLRKKLV DSTDKADLRL IYLALAHMIK  360
FRGHFLIEGD LNPDNSDVDK LFIQLVQTYN QLFEENPINA SGVDAKAILS ARLSKSRRLE  420
NLIAQLPGEK KNGLFGNLIA LSLGLTPNFK SNFDLAEDAK LQLSKDTYDD DLDNLLAQIG  480
DQYADLFLAA KNLSDAILLS DILRVNTEIT KAPLSASMIK RYDEHHQDLT LLKALVRQQL  540
PEKYKEIFFD QSKNGYAGYI DGGASQEEFY KFIKPILEKM DGTEELLVKL NREDLLRKQR  600
TFDNGSIPHQ IHLGELHAIL RRQEDFYPFL KDNREKIEKI LTFRIPYYVG PLARGNSRFA  660
WMTRKSEETI TPWNFEEVVD KGASAQSFIE RMTNFDKNLP NEKVLPKHSL LYEYFTVYNE  720
LTKVKYVTEG MRKPAFLSGE QKKAIVDLLF KTNRKVTVKQ LKEDYFKKIE CFDSVEISGV  780
EDRFNASLGT YHDLLKIIKD KDFLDNEENE DILEDIVLTL TLFEDREMIE ERLKTYAHLF  840
DDKVMKQLKR RRYTGWGRLS RKLINGIRDK QSGKTILDFL KSDGFANRNF MQLIHDDSLT  900
FKEDIQKAQV SGQGDSLHEH IANLAGSPAI KKGILQTVKV VDELVKVMGR HKPENIVIEM  960
ARENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY  1020
VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW  1080
RQLLNAKLIT QRKFDNLTKA ERGGLSELDK AGFIKRQLVE TRQITKHVAQ ILDSRMNTKY  1140
DENDKLIREV KVITLKSKLV SDFRKDFQFY KVREINNYHH AHDAYLNAVV GTALIKKYPK  1200
LESEFVYGDY KVYDVRKMIA KSEQEIGKAT AKYFFYSNIM NFFKTEITLA NGEIRKRPLI  1260
ETNGETGEIV WDKGRDFATV RKVLSMPQVN IVKKTEVQTG GFSKESILPK RNSDKLIARK  1320
KDWDPKKYGG FDSPTVAYSV LVVAKVEKGK SKKLKSVKEL LGITIMERSS FEKNPIDFLE  1380
AKGYKEVKKD LIIKLPKYSL FELENGRKRM LASAGELQKG NELALPSKYV NFLYLASHYE  1440
KLKGSPEDNE QKQLFVEQHK HYLDEIIEQI SEFSKRVILA DANLDKVLSA YNKHRDKPIR  1500
EQAENIIHLF TLTNLGAPAA FKYFDTTIDR KRYTSTKEVL DATLIHQSIT GLYETRIDLS  1560
QLGGD                                                              1565

SEQ ID NO: 82          moltype = AA  length = 364
FEATURE                Location/Qualifiers
source                 1..364
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 82
SEVEFSHEYW MRHALTLAKR AWDEREVPVG AVLVHNNRVI GEGWNRPIGR HDPTAHAEIM  60
ALRQGGLVMQ NYRLIDATLY VTLEPCVMCA GAMIHSRIGR VVFGARDAKT GAAGSLMDVL  120
HHPGMNHRVE ITEGILADEC AALLSDFFRM RRQEIKAQKK AQSSTDSGGS SGGSSGSETP  180
GTSESATPES SGGSSGGSSE VEFSHEYWMR HALTLAKRAR DEREVPVGAV LVLNNRVIGE  240
GWNRAIGLHD PTAHAEIMAL RQGGLVMQNY RLIDATLYVT FEPCVMCAGA MIHSRIGRVV  300
FGVRNAKTGA AGSLMDVLHY PGMNHRVEIT EGILADECAA LLCYFFRMPR QVFNAQKKAQ  360
SSTD                                                               364

SEQ ID NO: 83          moltype = AA  length = 167
FEATURE                Location/Qualifiers
source                 1..167
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 83
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI  60
MALRQGGLVM QNYRLYDATL YSTFEPCVMC AGAMIHSRIG RVVFGVRNAK TGAAGSLMDV  120
LHHPGMNHRV EITEGILADE CAALLCRFFR MPRRVFNAQK KAQSSTD                167

SEQ ID NO: 84          moltype = AA  length = 167
FEATURE                Location/Qualifiers
source                 1..167
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 84
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG LHDPTAHAEI  60
MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG RVVFGVRNSK RGAAGSLMNV  120
LNYPGMNHRV EITEGILADE CAALLCDFYR MPRQVFNAQK KAQSSIN                167

SEQ ID NO: 85          moltype = AA  length = 83
FEATURE                Location/Qualifiers
```

-continued

```
source                 1..83
                       mol_type = protein
                       note = Bacillus phage AR9
                       organism = unidentified
SEQUENCE: 85
TNLSDIIEKE TGKQLVIQES ILMLPEEVEE VIGNKPESDI LVHTAYDEST DENVMLLTSD  60
APEYKPWALV IQDSNGENKI KML                                         83

SEQ ID NO: 86          moltype = DNA  length = 66
FEATURE                Location/Qualifiers
source                 1..66
                       mol_type = other DNA
                       organism = Saccharomyces bayanus
SEQUENCE: 86
ttcttgtcgt acttatagat cgctacgtta tttcaatttt gaaaatctga gtcctgggag  60
tgcgga                                                            66

SEQ ID NO: 87          moltype = AA  length = 605
FEATURE                Location/Qualifiers
source                 1..605
                       mol_type = protein
                       organism = Homo sapiens
SEQUENCE: 87
MSGWESYYKT EGDEEAEEEQ EENLEASGDY KYSGRDSLIF LVDASKAMFE SQSEDELTPF  60
DMSIQCIQSV YISKIISSDR DLLAWFYGTE KDKNSVNFKI YVLQELDNPG AKRILELDQF  120
KGQQGQKRFQ DMMGHGSDYS LSEVLWVCAN LFSDVQFKMS HKRIMLFTNE DNPHGNDSAK  180
ASRARTKAGD LRDTGIFLDL HLKKPGGFDI SLFYRDIISI AEDEDLRVHF EESSKLEDLL  240
RKVRAKETRK RALSRLKLKL NKDIVISVGI YNLVQKALKP PPIKLYRETN EPVKTKTRTF  300
NTSTGGLLLP SDTKRSQIYG SRQIILEKEE TEELKRFDDP GLMLMGFKPL VLLKKHHYLR  360
PSLFVYPEES LVIGSSTLFS ALLIKCLEKE VAALCRYTPR RNIPPYFVAL VPQEEELDDQ  420
KIQVTPPGFQ LVFLPFADDK RKMPFTEKIM ATPEQVGKMK AIVEKLRFTY RSDSFENPVL  480
QQHFRNLEAL ALDLMEPEQA VDLTLPKVEA MNKRLGSLVD EFKELVYPPD YNPEGKVTKR  540
KHDNEGSGSK RPKVEYSEEE LKTHISKGTL GKFTVPLKEA CRAYGLKSGL KKQELLEALT  600
KHFQD                                                             605

SEQ ID NO: 88          moltype = AA  length = 482
FEATURE                Location/Qualifiers
REGION                 1..482
                       note = Synthetic polypeptide
source                 1..482
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 88
MVRSGNKAAW LCMDVGFTMS NSIPGIESPF EQAKKVITMF VQRQVFAENK DEIALVLFGT  60
DGTDNPLSGG DQYQNITVHR HLMLPDFDLL EDIESKIQPG SQQADFLDAL IVSMDVIQHE  120
TIGKKFEKRH IEIFTDLSSR FSKSQLDIII HSLKKCDISE RHSIHWPCRL TIGSNLSIRI  180
AAYKSILQER VKKTTWDAKT LKKEDIQKET VYCLNDDDET EVLKEDIIQG FRYGSDIVPF  240
SKVDEEQMKY KSEGKCFSVL GFCKSSQVQR RFFMGNQVLK VFAARDDEAA AVALSSLIHA  300
LDDLDIWAIV RYAYDKRANP QVGVAFPHIK HNYECLVYVQ LPFMEDLRQY MFSSLKNSKK  360
YAPTEAQLNA VDALIDSMSL AKKDEKTDTL EDLFPTTKIP NPRFQRLFQC LLHRALHPRE  420
PLPPIQQHIW NMLNPPAEVT TKSQIPLSKI KTLFPLIEAK KKDQVTAQEI FQDNHEDGPT  480
AK                                                                482

SEQ ID NO: 89          moltype = DNA  length = 10
FEATURE                Location/Qualifiers
source                 1..10
                       mol_type = other DNA
                       organism = Methanobacterium thermoautotrophicum
SEQUENCE: 89
aatttttgga                                                        10

SEQ ID NO: 90          moltype = AA  length = 83
FEATURE                Location/Qualifiers
source                 1..83
                       mol_type = protein
                       organism = Methanobacterium thermoautotrophicum
SEQUENCE: 90
GSVIDVSSQR VNVQRPLDAL GNSLNSPVII KLKGDREFRG VLKSFDLHMN LVLNDAEELE  60
DGEVTRRLGT VLIRGDNIVY ISP                                         83

SEQ ID NO: 91          moltype = DNA  length = 25
FEATURE                Location/Qualifiers
source                 1..25
                       mol_type = other DNA
                       organism = Emesvirus zinderi
SEQUENCE: 91
```

-continued

```
gcgcacatga ggatcaccca tgtgc                                       25

SEQ ID NO: 92          moltype = AA  length = 116
FEATURE                Location/Qualifiers
source                 1..116
                       mol_type = protein
                       organism = Emesvirus zinderi
SEQUENCE: 92
MASNFTQFVL VDNGGTGDVT VAPSNFANGI AEISSNSRSQ AYKVTCSVRQ SSAQNRKYTI  60
KVEVPKGAWR SYLNMELTIP IFATNSDCEL IVKAMQGLLK DGNPIPSAIA ANSGIY      116

SEQ ID NO: 93          moltype = DNA  length = 26
FEATURE                Location/Qualifiers
source                 1..26
                       mol_type = other DNA
                       organism = Pepevirus rubrum
SEQUENCE: 93
ataaggagtt tatatggaaa ccctta                                      26

SEQ ID NO: 94          moltype = AA  length = 127
FEATURE                Location/Qualifiers
source                 1..127
                       mol_type = protein
                       organism = Pepevirus rubrum
SEQUENCE: 94
MSKTIVLSVG EATRTLTEIQ STADRQIFEE KVGPLVGRLR LTASLRQNGA KTAYRVNLKL  60
DQADWDCSTS VCGELPKVRY TQVWSHDVTI VANSTEASRK SLYDLTKSLV ATSQVEDLVV  120
NLVPLGR                                                           127

SEQ ID NO: 95          moltype = DNA  length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = other DNA
                       organism = Shigella flexneri
SEQUENCE: 95
ctgaatgcct gcgagcatc                                              19

SEQ ID NO: 96          moltype = AA  length = 62
FEATURE                Location/Qualifiers
source                 1..62
                       mol_type = protein
                       organism = Shigella flexneri
SEQUENCE: 96
MKSIRCKNCN KLLFKADSFD HIEIRCPRCK RHIIMLNACE HPTEKHCGKR EKITHSDETV  60
RY                                                                62

SEQ ID NO: 97          moltype = AA  length = 24
FEATURE                Location/Qualifiers
source                 1..24
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 97
EELLSKNYHL ENEVARLKKG SGSG                                        24

SEQ ID NO: 98          moltype = AA  length = 241
FEATURE                Location/Qualifiers
source                 1..241
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 98
EEELLSKNYH LENEVARLKK GSGSGEELLS KNYHLENEVA RLKKGSGSGE ELLSKNYHLE  60
NEVARLKKGS GSGEELLSKN YHLENEVARL KKGSGSGEEL LSKNYHLENE VARLKKGSGS  120
GEELLSKNYH LENEVARLKK GSGSGEELLS KNYHLENEVA RLKKGSGSGE ELLSKNYHLE  180
NEVARLKKGS GSGEELLSKN YHLENEVARL KKGSGSGEEL LSKNYHLENE VARLKKGSGS  240
G                                                                 241

SEQ ID NO: 99          moltype = AA  length = 277
FEATURE                Location/Qualifiers
source                 1..277
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 99
MGPDIVMTQS PSSLSASVGD RVTITCRSST GAVTTSNYAS WVQEKPGKLF KGLIGGTNNR  60
APGVPSRFSG SLIGDKATLT ISSLQPEDFA TYFCALWYSN HWVFGQGTKV ELKRGGGGSG  120
GGGSGGGGSS GGGSEVKLLE SGGGLVQPGG SLKLSCAVSG FSLTDYGVNW VRQAPGRGLE  180
WIGVIWGDGI TDYNSALKDR FIISKDNGKN TVYLQMSKVR SDDTALYYCV TGLFDYWGQG  240
```

-continued

```
TLVTVSSYPY DVPDYAGGGG GSGGGGSGGG GSGGGGS                     277

SEQ ID NO: 100         moltype = AA  length = 1030
FEATURE                Location/Qualifiers
source                 1..1030
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 100
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPNSRL INSGGSSGGS SGSETPGTSE  720
SATPESSGGS SGGSSKLEKF TNCYSLSKTL RFKAIPVGKT QENIDNKRLL VEDEKRAEDY  780
KGVKKLLDRY YLSFINDVLH SIKLKNLNNY ISLFRKKTRT EKENKELENL EINLRKEIAK  840
AFKGNEGYKS LFKKDIIETI LPEFLDDKDE IALVNSFNGF TTAFTGFFDN RENMFSEEAK  900
STSIAFRCLS YETEILTVEY GLLPIGKIVE KRIECTVYSV DNNGNIYTQP VAQWHDRGEQ  960
EVFEYCLEDG SLIRATKDHK FMTVDGQMLP IDEIFERELD LMRVDNLPNS GGSKRTADGS  1020
EFEPKKKRKV                                                   1030

SEQ ID NO: 101         moltype = AA  length = 1197
FEATURE                Location/Qualifiers
source                 1..1197
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 101
MKRTADGSEF ESPKKKRKVI KIATRKYLGK QNVYDIGVGE PHNFALKNGF IASNCINENL  60
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  120
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  180
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  240
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  300
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  360
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  420
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  480
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  540
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  600
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS  660
YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY  720
IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK  780
AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK  840
KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS  900
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  960
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS  1020
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1080
AEDEKLDKVK IAISNKEWLE YAQTSVKHSG GSSGGSSGSE TPGTSESATP ESSGGSSGGS  1140
ALDFLSRLPL PPPVSPICTF VSPAAQKAFQ PPRSCGSGGS KRTADGSEFE PKKKRKV     1197

SEQ ID NO: 102         moltype = AA  length = 1166
FEATURE                Location/Qualifiers
source                 1..1166
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 102
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPNSRL INSGGSSGGS SGSETPGTSE  720
SATPESSGGS SGGSSKLEKF TNCYSLSKTL RFKAIPVGKT QENIDNKRLL VEDEKRAEDY  780
KGVKKLLDRY YLSFINDVLH SIKLKNLNNY ISLFRKKTRT EKENKELENL EINLRKEIAK  840
AFKGNEGYKS LFKKDIIETI LPEFLDDKDE IALVNSFNGF TTAFTGFFDN RENMFSEEAK  900
STSIAFRCIN ENLTRYISNM DIFEKVDAIF DKHEVQEIKE KILNSDYDVE DFFEGEFFNF  960
```

-continued

```
VLTQEGIDVY NAIIGGFVTE SGEKIKGLNE YINLYNQKTK QKLPKFKPLY KQVLSDRESL  1020
SFYGEGYTSD EEVLEVFRNT LNKCLSYETE ILTVEYGLLP IGKIVEKRIE CTVYSVDNNG  1080
NIYTQPVAQW HDRGEQEVFE YCLEDGSLIR ATKDHKFMTV DGQMLPIDEI FERELDLMRV  1140
DNLPNSGGSK RTADGSEFEP KKKRKV                                       1166

SEQ ID NO: 103         moltype = AA  length = 1063
FEATURE                Location/Qualifiers
source                 1..1063
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 103
MKRTADGSEF ESPKKKRKVI KIATRKYLGK QNVYDIGVGE PHNFALKNGF IASNCANSEI  60
FSSIKKLEKL FKNFDEYSSA GIFVKNGPAI STISKDIFGE WNVIRDKWNA EYDDIHLKKK  120
AVVTEKYEDD RRKSFKKIGS FSLEQLQEYA DADLSVVEKL KEIIIQKVDE IYKVYGSSEK  180
LFDADFVLEK SLKKNDAVVA IMKDLLDSVK SFENYIKAFF GEGKETNRDE SFYGDFVLAY  240
DILLKVDHIY DAIRNYVTQK PYSKDKFKLY FQNPQFMGGW DKDKETDYRA TILRYGSKYY  300
LAIMDKKYAK CLQKIDKDDV NGNYEKINYK LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK  360
IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD FNFSETEKYK DIAGFYREVE  420
EQGYKVSFES ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH GTPNLHTMYF KLLFDENNHG  480
QIRLSGGAEL FMRRASLKKE ELVVHPANSP IANKNPDNPK KTTTLSYDVY KDKRFSEDQY  540
ELHIPIAINK CPKNIFKINT EVRVLLKHDD NPYVIGIDRG ERNLLYIVVV DGKGNIVEQY  600
SLNEIINNFN GIRIKTDYHS LLDKKEKERF EARQNWTSIE NIKELKAGYI SQVVHKICEL  660
VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL NYMVDKKSNP CATGGALKGY  720
QITNKFESFK SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL KTKYTSIADS KKFISSFDRI  780
MYVPEEDLFE FALDYKNFSR TDADYIKKWK LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA  840
YKELFNKYGI NYQQGDIRAL LCEQSDKAFY SSFMALMSLM LQMRNSITGR TDVDFLISPV  900
KNSDGIFYDS RNYEAQENAI LPKNADANGA YNIARKVLWA IGQFKKAEDE KLDKVKIAIS  960
NKEWLEYAQT SVKHSGGSSG GSSGSETPGT SESATPESSG GSSGGSALDF LSRLPLPPPV  1020
SPICTFVSPA AQKAFQPPRS CGSGGSKRTA DGSEFEPKKK RKV                    1063

SEQ ID NO: 104         moltype = AA  length = 1262
FEATURE                Location/Qualifiers
source                 1..1262
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 104
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPNSRL INSGGSSGGS SGSETPGTSE  720
SATPESSGGS SGGSSKLEKF TNCYSLSKTL RFKAIPVGKT QENIDNKRLL VEDEKRAEDY  780
KGVKKLLDRY YLSFINDVLH SIKLKNLNNY ISLFRKKTRT EKENKELENL EINLRKEIAK  840
AFKGNEGYKS LFKKDIIETI LPEFLDDKDE IALVNSFNGF TTAFTGFFDN RENMFSEEAK  900
STSIAFRCIN ENLTRYISNM DIFEKVDAIF DKHEVQEIKE KILNSDYDVE DFFEGEFFNF  960
VLTQEGIDVY NAIIGGFVTE SGEKIKGLNE YINLYNQKTK QKLPKFKPLY KQVLSDRESL  1020
SFYGEGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI  1080
SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAC  1140
LSYETEILTV EYGLLPIGKI VEKRIECTVY SVDNNGNIYT QPVAQWHDRG EQEVFEYCLE  1200
DGSLIRATKD HKFMTVDGQM LPIDEIFERE LDLMRVDNLP NSGGSKRTAD GSEFEPKKKR  1260
KV                                                                 1262

SEQ ID NO: 105         moltype = AA  length = 967
FEATURE                Location/Qualifiers
source                 1..967
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 105
MKRTADGSEF ESPKKKRKVI KIATRKYLGK QNVYDIGVGE PHNFALKNGF IASNCADLSV  60
VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI  120
KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF  180
MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN  240
KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS  300
NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS  360
DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP  420
DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG  480
IDRGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW  540
TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML  600
IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF  660
```

-continued

```
VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN  720
RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL  780
MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK  840
VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKHSG GSSGGSSGSE TPGTSESATP  900
ESSGGSSGGS ALDFLSRLPL PPPVSPICTF VSPAAQKAFQ PPRSCGSGGS KRTADGSEFE  960
PKKKRKV                                                            967

SEQ ID NO: 106          moltype = AA  length = 1297
FEATURE                 Location/Qualifiers
source                  1..1297
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 106
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPNSRL INSGGSSGGS SGSETPGTSE  720
SATPESSGGS SGGSSKLEKF TNCYSLSKTL RFKAIPVGKT QENIDNKRLL VEDEKRAEDY  780
KGVKKLLDRY YLSFINDVLH SIKLKNLNNY ISLFRKKTRT EKENKELENL EINLRKEIAK  840
AFKGNEGYKS LFKKDIIETI LPEFLDDKDE IALVNSFNGF TTAFTGFFDN RENMFSEEAK  900
STSIAFRCIN ENLTRYISNM DIFEKVDAIF DKHEVQEIKE KILNSDYDVE DFFEGEFFNF  960
VLTQEGIDVY NAIIGGFVTE SGEKIKGLNE YINLYNQKTK QKLPKFKPLY KQVLSDRESL  1020
SFYGEGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI  1080
SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAD  1140
LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVCLSYET EILTVEYGLL PIGKIVEKRI  1200
ECTVYSVDNN GNIYTQPVAQ WHDRGEQEVF EYCLEDGSLI RATKDHKFMT VDGQMLPIDE  1260
IFERELDLMR VDNLPNSGGS KRTADGSEFE PKKKRKV                           1297

SEQ ID NO: 107          moltype = AA  length = 932
FEATURE                 Location/Qualifiers
source                  1..932
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 107
MKRTADGSEF ESPKKKRKVI KIATRKYLGK QNVYDIGVGE PHNFALKNGF IASNCALEKS  60
LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD  120
AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC  180
LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD  240
MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA  300
SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF  360
MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC  420
PKNIFKINTE VRVLLKHDDN PYVIGIDRGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG  480
IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE  540
DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS  600
MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF  660
ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN  720
YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR  780
NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS  840
VKHSGGSSGG SSGSETPGTS ESATPESSGG SSGGSALDFL SRLPLPPPVS PICTFVSPAA  900
QKAFQPPRSC GSGGSKRTAD GSEFEPKKKR KV                                932

SEQ ID NO: 108          moltype = AA  length = 1406
FEATURE                 Location/Qualifiers
source                  1..1406
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 108
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
```

-continued

```
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPNSRL INSGGSSGGS SGSETPGTSE  720
SATPESSGGS SGGSSKLEKF TNCYSLSKTL RFKAIPVGKT QENIDNKRLL VEDEKRAEDY  780
KGVKKLLDRY YLSFINDVLH SIKLKNLNNY ISLFRKKTRT EKENKELENL EINLRKEIAK  840
AFKGNEGYKS LFKKDIIETI LPEFLDDKDE IALVNSFNGF TTAFTGFFDN RENMFSEEAK  900
STSIAFRCIN ENLTRYISNM DIFEKVDAIF DKHEVQEIKE KILNSDYDVE DFFEGEFFNF  960
VLTQEGIDVY NAIIGGFVTE SGEKIKGLNE YINLYNQKTK QKLPKFKPLY KQVLSDRESL  1020
SFYGEGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI  1080
SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAD  1140
LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK KNDAVVAIMK DLLDSVKSFE  1200
NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI RNYVTQKPYS KDKFKLYFQN  1260
PQFMGGWDKD KETDYRATIL RYGCLSYETE ILTVEYGLLP IGKIVEKRIE CTVYSVDNNG  1320
NIYTQPVAQW HDRGEQEVFE YCLEDGSLIR ATKDHKFMTV DGQMLPIDEI FERELDLMRV  1380
DNLPNSGGSK RTADGSEFEP KKKRKV                                       1406

SEQ ID NO: 109          moltype = AA  length = 823
FEATURE                 Location/Qualifiers
source                  1..823
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 109
MKRTADGSEF ESPKKKRKVI KIATRKYLGK QNVYDIGVGE PHNFALKNGF IASNCASKYY  60
LAIMDKKYAK CLQKIDKDDV NGNYEKINYK LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK  120
IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD FNFSETEKYK DIAGFYREVE  180
EQGYKVSFES ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH GTPNLHTMYF KLLFDENNHG  240
QIRLSGGAEL FMRRASLKKE ELVVHPANSP IANKNPDNPK KTTTLSYDVY KDKRFSEDQY  300
ELHIPIAINK CPKNIFKINT EVRVLLKHDD NPYVIGIDRG ERNLLYIVVV DGKGNIVEQY  360
SLNEIINNFN GIRIKTDYHS LLDKKEKERF EARQNWTSIE NIKELKAGYI SQVVHKICEL  420
VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL NYMVDKKSNP CATGGALKGY  480
QITNKFESFK SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL KTKYTSIADS KKFISSFDRI  540
MYVPEEDLFE FALDYKNFSR TDADYIKKWK LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA  600
YKELFNKYGI NYQQGDIRAL LCEQSDKAFY SSFMALMSLM LQMRNSITGR TDVDFLISPV  660
KNSDGIFYDS RNYEAQENAI LPKNADANGA YNIARKVLWA IGQFKKAEDE KLDKVKIAIS  720
NKEWLEYAQT SVKHSGGSSG GSSGSETPGT SESATPESSG GSSGGSALDF LSRLPLPPPV  780
SPICTFVSPA AQKAFQPPRS CGSGGSKRTA DGSEFEPKKK RKV                    823

SEQ ID NO: 110          moltype = AA  length = 102
FEATURE                 Location/Qualifiers
source                  1..102
                        mol_type = protein
                        organism = Nostoc punctiforme
SEQUENCE: 110
CLSYETEILT VEYGLLPIGK IVEKRIECTV YSVDNNGNIY TQPVAQWHDR GEQEVFEYCL  60
EDGSLIRATK DHKFMTVDGQ MLPIDEIFER ELDLMRVDNL PN                     102

SEQ ID NO: 111          moltype = AA  length = 35
FEATURE                 Location/Qualifiers
source                  1..35
                        mol_type = protein
                        organism = Nostoc punctiforme
SEQUENCE: 111
IKIATRKYLG KQNVYDIGVG EPHNFALKNG FIASN                             35

SEQ ID NO: 112          moltype = AA  length = 35
FEATURE                 Location/Qualifiers
source                  1..35
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 112
IKIATRKYLG KQNVYDIGVG EPHNFALKNG FIASN                             35

SEQ ID NO: 113          moltype = AA  length = 2010
FEATURE                 Location/Qualifiers
source                  1..2010
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 113
TLNIEDEHRL HETSKEPDVS LGSTWLSDFP QAWAETGGMG LAVRQAPLII PLKATSTPVS  60
IKQYPMSQEA RLGIKPHIQR LLDQGILVPC QSPWNTPLLP VKKPGTNDYR PVQDLREVNK  120
RVEDIHPTVP NPYNLLSGLP PSHQWYTVLD LKDAFFCLRL HPTSQPLFAF EWRDPEMGIS  180
GQLTWTRLPQ GFKNSPTLFN EALHRDLADF RIQHPDLILL QYVDDLLLAA TSELDCQQGT  240
RALLQTLGNL GYRASAKKAQ ICQKQVKYLG YLLKEGQRWL TEARKETVMG QPTPKTPRQL  300
REFLGKAGFC RLFIPGFAEM AAPLYPLTKP GTLFNWGPDQ QKAYQEIKQA LLTAPALGLP  360
DLTKPFELFV DEKQGYAKGV LTQKLGPWRR PVAYLSKKLD PVAAGWPPCL RMVAAIAVLT  420
KDAGKLTMGQ PLVILAPHAV EALVKQPPDR WLSNARMTHY QALLLDTDRV QFGPVVALNP  480
ATLLPLPEEG LQHNCLDILA EAHGTRPDLT DQPLPDADHT WYTDGSSLLQ EGQRKAGAAV  540
TTETEVIWAK ALPAGTSAQR AELIALTQAL KMAEGKKLNV YTDSRYAFAT AHIHGEIYRR  600
```

-continued

```
RGWLTSEGKE IKNKDEILAL LKALFLPKRL SIIHCPGHQK GHSAEARGNR MADQAARKAA  660
ITETPDTSTL LIENSSPNSR LINSGGSSGG SSGSETPGTS ESATPESSGG SSGGSSKLEK  720
FTNCYSLSKT LRFKAIPVGK TQENIDNKRL LVEDEKRAED YKGVKKLLDR YYLSFINDVL  780
HSIKLKNLNN YISLFRKKTR TEKENKELEN LEINLRKEIA KAFKGNEGYK SLFKKDIIET  840
ILPEFLDDKD EIALVNSFNG FTTAFTGFFD NRENMFSEEA KSTSIAFRCI NENLTRYISN  900
MDIFEKVDAI FDKHEVQEIK EKILNSDYDV EDFFEGEFFN FVLTQEGIDV YNAIIGGFVT  960
ESGEKIKGLN EYINLYNQKT KQKLPKFKPL YKQVLSDRES LSFYGEGYTS DEEVLEVFRN  1020
TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI FVKNGPAIST ISKDIFGEWN VIRDKWNAEY  1080
DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS LEQLQEYADA DLSVVEKLKE IIIQKVDEIY  1140
KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE GKETNRDESF  1200
YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK DKETDYRATI  1260
LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL PGPNKMLPKV FFSKKWMAYY  1320
NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY PKWSNAYDFN FSETEKYKDI  1380
AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL  1440
LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA NKNPDNPKKT TTLSYDVYKD  1500
KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP YVIGIDRGER NLLYIVVVDG  1560
KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA RQNWTSIENI KELKAGYISQ  1620
VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA  1680
TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP STGFVNLLKT KYTSIADSKK  1740
FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW  1800
EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS FMALMSLMLQ MRNSITGRTD  1860
VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN IARKVLWAIG QFKKAEDEKL  1920
DKVKIAISNK EWLEYAQTSV KHSGGSSGGS SGSETPGTSE SATPESSGGS SGGSALDFLS  1980
RLPLPPPVSP ICTFVSPAAQ KAFQPPRSCG                                  2010

SEQ ID NO: 114          moltype = AA  length = 1373
FEATURE                 Location/Qualifiers
source                  1..1373
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 114
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN  720
GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL  780
IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE  840
EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE  900
LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE  960
VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL  1020
LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR  1080
VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY  1140
IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT  1200
DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL  1260
CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL  1320
PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH         1373

SEQ ID NO: 115          moltype = AA  length = 1375
FEATURE                 Location/Qualifiers
source                  1..1375
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 115
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGSGENQTTQ  300
KGQKNSRERM KRIEEGIKEL GSQILKEHPV ENTQLQNEKL YLYYLQNGRD MYVDQELDIN  360
RLSDYDVDHI VPQSFLKDDS IDNKVLTRSD KNRGKSDNVP SEEVVKKMKN YWRQLLNAKL  420
ITQRKFDNLT KAERGGLSEG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
```

-continued

```
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1375

SEQ ID NO: 116          moltype = AA  length = 1377
FEATURE                 Location/Qualifiers
source                  1..1377
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 116
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGGSSGENQT  300
TQKGQKNSRE RMKRIEEGIK ELGSQILKEH PVENTQLQNE KLYLYYLQNG RDMYVDQELD  360
INRLSDYDVD HIVPQSFLKD DSIDNKVLTR SDKNRGKSDN VPSEEVVKKM KNYWRQLLNA  420
KLITQRKFDN LTKAERGGLS EGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  480
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  540
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  600
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  660
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  720
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  780
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  840
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  900
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  960
INTEVRVLLK HDDNPYVIGI ARGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  1020
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1080
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1140
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1200
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1260
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1320
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1377

SEQ ID NO: 117          moltype = AA  length = 1379
FEATURE                 Location/Qualifiers
source                  1..1379
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 117
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGGSSGENQT  300
TQKGQKNSRE RMKRIEEGIK ELGSQILKEH PVENTQLQNE KLYLYYLQNG RDMYVDQELD  360
INRLSDYDVD HIVPQSFLKD DSIDNKVLTR SDKNRGKSDN VPSEEVVKKM KNYWRQLLNA  420
KLITQRKFDN LTKAERGGLS GSEGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD  480
EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF  540
KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN  600
DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN  660
YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI  720
DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL  780
NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE  840
VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA  900
SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI  960
FKINTEVRVL LKHDDNPYVI GIARGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK  1020
TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS  1080
GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ  1140
NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY  1200
KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG  1260
DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA  1320
QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH   1379

SEQ ID NO: 118          moltype = AA  length = 1373
FEATURE                 Location/Qualifiers
source                  1..1373
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 118
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
```

-continued

```
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEENQTTQK  300
GQKNSRERMK RIEEGIKELG SQILKEHPVE NTQLQNEKLY LYYLQNGRDM YVDQELDINR  360
LSDYDVDHIV PQSFLKDDSI DNKVLTRSDK NRGKSDNVPS EEVVKKMKNY WRQLLNAKLI  420
TQRKFDNLTK AERGGLSGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN  720
GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL  780
IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE  840
EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE  900
LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE  960
VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL  1020
LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR  1080
VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY  1140
IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT  1200
DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL  1260
CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL  1320
PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH         1373

SEQ ID NO: 119         moltype = AA  length = 1375
FEATURE                Location/Qualifiers
source                 1..1375
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 119
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSENQTT  300
QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI  360
NRLSDYDVDH IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK  420
LITQRKFDNL TKAERGGLSG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1375

SEQ ID NO: 120         moltype = AA  length = 1377
FEATURE                Location/Qualifiers
source                 1..1377
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 120
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSSGENQ  300
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  360
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  420
AKLITQRKFD NLTKAERGGL SGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  480
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  540
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  600
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  660
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  720
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  780
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  840
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  900
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  960
INTEVRVLLK HDDNPYVIGI ARGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  1020
```

-continued

```
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1080
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1140
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1200
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1260
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1320
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1377

SEQ ID NO: 121          moltype = AA  length = 1379
FEATURE                 Location/Qualifiers
source                  1..1379
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 121
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSSGENQ  300
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  360
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  420
AKLITQRKFD NLTKAERGGL SGSGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD  480
EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF  540
KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN  600
DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN  660
YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI  720
DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL  780
NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE  840
VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA  900
SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI  960
FKINTEVRVL LKHDDNPYVI GIARGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK  1020
TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS  1080
GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ  1140
NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY  1200
KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG  1260
DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA  1320
QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH   1379

SEQ ID NO: 122          moltype = AA  length = 1373
FEATURE                 Location/Qualifiers
source                  1..1373
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 122
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN  720
GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL  780
IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE  840
EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE  900
LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE  960
VRVLLKHDDN PYVIGIDRGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL  1020
LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR  1080
VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY  1140
IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT  1200
DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL  1260
CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL  1320
PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH         1373

SEQ ID NO: 123          moltype = AA  length = 1375
FEATURE                 Location/Qualifiers
source                  1..1375
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 123
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
```

-continued

```
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGSGENQTTQ  300
KGQKNSRERM KRIEEGIKEL GSQILKEHPV ENTQLQNEKL YLYYLQNGRD MYVDQELDIN  360
RLSDYDVDHI VPQSFLKDDS IDNKVLTRSD KNRGKSDNVP SEEVVKKMKN YWRQLLNAKL  420
ITQRKFDNLT KAERGGLSEG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIDR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1375

SEQ ID NO: 124         moltype = AA  length = 1377
FEATURE                Location/Qualifiers
source                 1..1377
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 124
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGGSSGENQT  300
TQKGQKNSRE RMKRIEEGIK ELGSQILKEH PVENTQLQNE KLYLYYLQNG RDMYVDQELD  360
INRLSDYDVD HIVPQSFLKD DSIDNKVLTR SDKNRGKSDN VPSEEVVKKM KNYWRQLLNA  420
KLITQRKFDN LTKAERGGLS EGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  480
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  540
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  600
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  660
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  720
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  780
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  840
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  900
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  960
INTEVRVLLK HDDNPYVIGI DRGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  1020
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1080
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1140
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1200
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1260
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1320
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1377

SEQ ID NO: 125         moltype = AA  length = 1379
FEATURE                Location/Qualifiers
source                 1..1379
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 125
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGGSSGENQT  300
TQKGQKNSRE RMKRIEEGIK ELGSQILKEH PVENTQLQNE KLYLYYLQNG RDMYVDQELD  360
INRLSDYDVD HIVPQSFLKD DSIDNKVLTR SDKNRGKSDN VPSEEVVKKM KNYWRQLLNA  420
KLITQRKFDN LTKAERGGLS GSEGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD  480
EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF  540
KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN  600
DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN  660
YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI  720
DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL  780
NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE  840
VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA  900
SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI  960
FKINTEVRVL LKHDDNPYVI GIDRGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK  1020
TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS  1080
GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ  1140
```

-continued

```
NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY  1200
KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG  1260
DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA  1320
QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH   1379

SEQ ID NO: 126          moltype = AA  length = 1373
FEATURE                 Location/Qualifiers
source                  1..1373
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 126
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEENQTTQK  300
GQKNSRERMK RIEEGIKELG SQILKEHPVE NTQLQNEKLY LYYLQNGRDM YVDQELDINR  360
LSDYDVDHIV PQSFLKDDSI DNKVLTRSDK NRGKSDNVPS EEVVKKMKNY WRQLLNAKLI  420
TQRKFDNLTK AERGGLSGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN  720
GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL  780
IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE  840
EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE  900
LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE  960
VRVLLKHDDN PYVIGIDRGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL  1020
LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR  1080
VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY  1140
IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT  1200
DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL  1260
CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL  1320
PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH         1373

SEQ ID NO: 127          moltype = AA  length = 1375
FEATURE                 Location/Qualifiers
source                  1..1375
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 127
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSENQTT  300
QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI  360
NRLSDYDVDH IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK  420
LITQRKFDNL TKAERGGLSG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIDR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1375

SEQ ID NO: 128          moltype = AA  length = 1377
FEATURE                 Location/Qualifiers
source                  1..1377
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 128
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSSGENQ  300
```

-continued

```
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  360
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  420
AKLITQRKFD NLTKAERGGL SGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  480
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  540
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  600
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  660
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  720
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  780
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  840
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  900
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  960
INTEVRVLLK HDDNPYVIGI DRGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  1020
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1080
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1140
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1200
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1260
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1320
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1377

SEQ ID NO: 129         moltype = AA  length = 1379
FEATURE                Location/Qualifiers
source                 1..1379
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 129
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSSGENQ  300
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  360
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  420
AKLITQRKFD NLTKAERGGL SGSGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD  480
EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF  540
KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN  600
DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN  660
YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI  720
DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL  780
NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE  840
VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA  900
SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI  960
FKINTEVRVL LKHDDNPYVI GIDRGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK  1020
TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS  1080
GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ  1140
NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY  1200
KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG  1260
DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA  1320
QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH   1379

SEQ ID NO: 130         moltype = AA  length = 1373
FEATURE                Location/Qualifiers
source                 1..1373
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 130
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN  720
GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL  780
IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE  840
EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE  900
LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE  960
VRVLLKHDDN PYVIGIDRGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL  1020
LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR  1080
VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY  1140
IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT  1200
DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL  1260
```

-continued

```
CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL  1320
PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH         1373

SEQ ID NO: 131          moltype = AA  length = 1375
FEATURE                 Location/Qualifiers
source                  1..1375
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 131
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGSGENQTTQ  300
KGQKNSRERM KRIEEGIKEL GSQILKEHPV ENTQLQNEKL YLYYLQNGRD MYVDQELDIN  360
RLSDYDVDHI VPQSFLKDDS IDNKVLTRSD KNRGKSDNVP SEEVVKKMKN YWRQLLNAKL  420
ITQRKFDNLT KAERGGLSEG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIDR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1375

SEQ ID NO: 132          moltype = AA  length = 1377
FEATURE                 Location/Qualifiers
source                  1..1377
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 132
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGGSSGENQT  300
TQKGQKNSRE RMKRIEEGIK ELGSQILKEH PVENTQLQNE KLYLYYLQNG RDMYVDQELD  360
INRLSDYDVD HIVPQSFLKD DSIDNKVLTR SDKNRGKSDN VPSEEVVKKM KNYWRQLLNA  420
KLITQRKFDN LTKAERGGLS EGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  480
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  540
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  600
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  660
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  720
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  780
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  840
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  900
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  960
INTEVRVLLK HDDNPYVIGI DRGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  1020
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1080
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1140
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1200
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1260
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1320
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1377

SEQ ID NO: 133          moltype = AA  length = 1379
FEATURE                 Location/Qualifiers
source                  1..1379
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 133
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGGSSGENQT  300
TQKGQKNSRE RMKRIEEGIK ELGSQILKEH PVENTQLQNE KLYLYYLQNG RDMYVDQELD  360
INRLSDYDVD HIVPQSFLKD DSIDNKVLTR SDKNRGKSDN VPSEEVVKKM KNYWRQLLNA  420
```

-continued

```
KLITQRKFDN LTKAERGGLS GSEGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD  480
EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF  540
KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN  600
DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN  660
YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI  720
DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL  780
NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE  840
VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA  900
SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI  960
FKINTEVRVL LKHDDNPYVI GIDRGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK  1020
TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS  1080
GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ  1140
NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY  1200
KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG  1260
DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA  1320
QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH   1379
```

```
SEQ ID NO: 134         moltype = AA  length = 1373
FEATURE                Location/Qualifiers
source                 1..1373
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 134
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEENQTTQK  300
GQKNSRERMK RIEEGIKELG SQILKEHPVE NTQLQNEKLY LYYLQNGRDM YVDQELDINR  360
LSDYDVDHIV PQSFLKDDSI DNKVLTRSDK NRGKSDNVPS EEVVKKMKNY WRQLLNAKLI  420
TQRKFDNLTK AERGGLSGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN  720
GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL  780
IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE  840
EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE  900
LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE  960
VRVLLKHDDN PYVIGIDRGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL  1020
LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR  1080
VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY  1140
IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT  1200
DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL  1260
CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL  1320
PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH         1373
```

```
SEQ ID NO: 135         moltype = AA  length = 1375
FEATURE                Location/Qualifiers
source                 1..1375
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 135
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSENQTT  300
QKGQKNSRER MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI  360
NRLSDYDVDH IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK  420
LITQRKFDNL TKAERGGLSG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIDR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1375
```

-continued

```
SEQ ID NO: 136         moltype = AA  length = 1377
FEATURE                Location/Qualifiers
source                 1..1377
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 136
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSSGENQ  300
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  360
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  420
AKLITQRKFD NLTKAERGGL SGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  480
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  540
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  600
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  660
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  720
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  780
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  840
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  900
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  960
INTEVRVLLK HDDNPYVIGI DRGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  1020
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1080
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1140
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1200
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1260
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1320
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1377

SEQ ID NO: 137         moltype = AA  length = 1379
FEATURE                Location/Qualifiers
source                 1..1379
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 137
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGSSGENQ  300
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  360
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  420
AKLITQRKFD NLTKAERGGL SGSGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD  480
EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF  540
KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN  600
DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN  660
YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI  720
DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL  780
NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE  840
VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA  900
SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI  960
FKINTEVRVL LKHDDNPYVI GIDRGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK  1020
TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS  1080
GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ  1140
NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY  1200
KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG  1260
DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA  1320
QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH   1379

SEQ ID NO: 138         moltype = AA  length = 1787
FEATURE                Location/Qualifiers
REGION                 1..1787
                       note = Synthetic polypeptide
source                 1..1787
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 138
MKRTADGSEF ESPKKKRKVS KLEKFTNCYS LSKTLRFKAI PVGKTQENID NKRLLVEDEK  60
RAEDYKGVKK LLDRYYLSFI NDVLHSIKLK NLNNYISLFR KKTRTEKENK ELENLEINLR  120
KEIAKAFKGN EGYKSLFKKD IIETILPEFL DDKDEIALVN SFNGFTTAFT GFFDNRENMF  180
SEEAKSTSIA FRCINENLTR YISNMDIFEK VDAIFDKHEV QEIKEKILNS DYDVEDFFEG  240
EFFNFVLTQE GIDVYNAIIG GFVTESGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS  300
DRESLSFYGE GSSGENQTTQ KGQKNSRERM KRIEEGIKEL GSQILKEHPV ENTQLQNEKL  360
YLYYLQNGRD MYVDQELDIN RLSDYDVDHI VPQSFLKDDS IDNKVLTRSD KNRGKSDNVP  420
```

-continued

```
SEEVVKKMKN YWRQLLNAKL ITQRKFDNLT KAERGGLSGY TSDEEVLEVF RNTLNKNSEI  480
FSSIKKLEKL FKNFDEYSSA GIFVKNGPAI STISKDIFGE WNVIRDKWNA EYDDIHLKKK  540
AVVTEKYEDD RRKSFKKIGS FSLEQLQEYA DADLSVVEKL KEIIIQKVDE IYKVYGSSEK  600
LFDADFVLEK SLKKNDAVVA IMKDLLDSVK SFENYIKAFF GEGKETNRDE SFYGDFVLAY  660
DILLKVDHIY DAIRNYVTQK PYSKDKFKLY FQNPQFMGGW DKDKETDYRA TILRYGSKYY  720
LAIMDKKYAK CLQKIDKDDV NGNYEKINYK LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK  780
IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD FNFSETEKYK DIAGFYREVE  840
EQGYKVSFES ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH GTPNLHTMYF KLLFDENNHG  900
QIRLSGGAEL FMRRASLKKE ELVVHPANSP IANKNPDNPK KTTTLSYDVY KDKRFSEDQY  960
ELHIPIAINK CPKNIFKINT EVRVLLKHDD NPYVIGIARG ERNLLYIVVV DGKGNIVEQY  1020
SLNEIINNFN GIRIKTDYHS LLDKKEKERF EARQNWTSIE NIKELKAGYI SQVVHKICEL  1080
VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL NYMVDKKSNP CATGGALKGY  1140
QITNKFESFK SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL KTKYTSIADS KKFISSFDRI  1200
MYVPEEDLFE FALDYKNFSR TDADYIKKWK LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA  1260
YKELFNKYGI NYQQGDIRAL LCEQSDKAFY SSFMALMSLM LQMRNSITGR TDVDFLISPV  1320
KNSDGIFYDS RNYEAQENAI LPKNADANGA YNIARKVLWA IGQFKKAEDE KLDKVKIAIS  1380
NKEWLEYAQT SVKHGGGGSG GGGSGGGGSG PKKKRKVAAA GSEELLSKNY HLENEVARLK  1440
KGSGSGGSGS GGSGSGSGGS GSGGSGSGEE LLSKNYHLEN EVARLKKGSG SGGSGSGGSG  1500
SGSGGSGSGG SGSGEELLSK NYHLENEVAR LKKGSGSGGS GSGGSGSGSG GSGSGGSGSG  1560
EELLSKNYHL ENEVARLKKG SGSGGSGSGG SGSGSGGSGS GGSGSGEELL SKNYHLENEV  1620
ARLKKGSGSG GSGSGGSGSG SGGSGSGGSG SGEELLSKNY HLENEVARLK KGSGSGGSGS  1680
GGSGSGSGGS GSGGSGSGEE LLSKNYHLEN EVARLKKGSG SGGSGSGGSG SGSGGSGSGG  1740
SGSGEELLSK NYHLENEVAR LKKSGGSKRT ADGSEFEPKK KRKVGSG              1787

SEQ ID NO: 139          moltype = AA  length = 1787
FEATURE                 Location/Qualifiers
REGION                  1..1787
                        note = Synthetic polypeptide
source                  1..1787
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 139
MKRTADGSEF ESPKKKRKVS KLEKFTNCYS LSKTLRFKAI PVGKTQENID NKRLLVEDEK  60
RAEDYKGVKK LLDRYYLSFI NDVLHSIKLK NLNNYISLFR KKTRTEKENK ELENLEINLR  120
KEIAKAFKGN EGYKSLFKKD IIETILPEFL DDKDEIALVN SFNGFTTAFT GFFDNRENMF  180
SEEAKSTSIA FRCINENLTR YISNMDIFEK VDAIFDKHEV QEIKEKILNS DYDVEDFFEG  240
EFFNFVLTQE GIDVYNAIIG GFVTESGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS  300
DRESLSFYGG SSGENQTTQK GQKNSRERMK RIEEGIKELG SQILKEHPVE NTQLQNEKLY  360
LYYLQNGRDM YVDQELDINR LSDYDVDHIV PQSFLKDDSI DNKVLTRSDK NRGKSDNVPS  420
EEVVKKMKNY WRQLLNAKLI TQRKFDNLTK AERGGLSEGY TSDEEVLEVF RNTLNKNSEI  480
FSSIKKLEKL FKNFDEYSSA GIFVKNGPAI STISKDIFGE WNVIRDKWNA EYDDIHLKKK  540
AVVTEKYEDD RRKSFKKIGS FSLEQLQEYA DADLSVVEKL KEIIIQKVDE IYKVYGSSEK  600
LFDADFVLEK SLKKNDAVVA IMKDLLDSVK SFENYIKAFF GEGKETNRDE SFYGDFVLAY  660
DILLKVDHIY DAIRNYVTQK PYSKDKFKLY FQNPQFMGGW DKDKETDYRA TILRYGSKYY  720
LAIMDKKYAK CLQKIDKDDV NGNYEKINYK LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK  780
IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD FNFSETEKYK DIAGFYREVE  840
EQGYKVSFES ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH GTPNLHTMYF KLLFDENNHG  900
QIRLSGGAEL FMRRASLKKE ELVVHPANSP IANKNPDNPK KTTTLSYDVY KDKRFSEDQY  960
ELHIPIAINK CPKNIFKINT EVRVLLKHDD NPYVIGIARG ERNLLYIVVV DGKGNIVEQY  1020
SLNEIINNFN GIRIKTDYHS LLDKKEKERF EARQNWTSIE NIKELKAGYI SQVVHKICEL  1080
VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL NYMVDKKSNP CATGGALKGY  1140
QITNKFESFK SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL KTKYTSIADS KKFISSFDRI  1200
MYVPEEDLFE FALDYKNFSR TDADYIKKWK LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA  1260
YKELFNKYGI NYQQGDIRAL LCEQSDKAFY SSFMALMSLM LQMRNSITGR TDVDFLISPV  1320
KNSDGIFYDS RNYEAQENAI LPKNADANGA YNIARKVLWA IGQFKKAEDE KLDKVKIAIS  1380
NKEWLEYAQT SVKHGGGGSG GGGSGGGGSG PKKKRKVAAA GSEELLSKNY HLENEVARLK  1440
KGSGSGGSGS GGSGSGSGGS GSGGSGSGEE LLSKNYHLEN EVARLKKGSG SGGSGSGGSG  1500
SGSGGSGSGG SGSGEELLSK NYHLENEVAR LKKGSGSGGS GSGGSGSGSG GSGSGGSGSG  1560
EELLSKNYHL ENEVARLKKG SGSGGSGSGG SGSGSGGSGS GGSGSGEELL SKNYHLENEV  1620
ARLKKGSGSG GSGSGGSGSG SGGSGSGGSG SGEELLSKNY HLENEVARLK KGSGSGGSGS  1680
GGSGSGSGGS GSGGSGSGEE LLSKNYHLEN EVARLKKGSG SGGSGSGGSG SGSGGSGSGG  1740
SGSGEELLSK NYHLENEVAR LKKSGGSKRT ADGSEFEPKK KRKVGSG              1787

SEQ ID NO: 140          moltype = AA  length = 1789
FEATURE                 Location/Qualifiers
REGION                  1..1789
                        note = Synthetic polypeptide
source                  1..1789
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 140
MKRTADGSEF ESPKKKRKVS KLEKFTNCYS LSKTLRFKAI PVGKTQENID NKRLLVEDEK  60
RAEDYKGVKK LLDRYYLSFI NDVLHSIKLK NLNNYISLFR KKTRTEKENK ELENLEINLR  120
KEIAKAFKGN EGYKSLFKKD IIETILPEFL DDKDEIALVN SFNGFTTAFT GFFDNRENMF  180
SEEAKSTSIA FRCINENLTR YISNMDIFEK VDAIFDKHEV QEIKEKILNS DYDVEDFFEG  240
EFFNFVLTQE GIDVYNAIIG GFVTESGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS  300
```

-continued

```
DRESLSFYGG SSGENQTTQK GQKNSRERMK RIEEGIKELG SQILKEHPVE NTQLQNEKLY  360
LYYLQNGRDM YVDQELDINR LSDYDVDHIV PQSFLKDDSI DNKVLTRSDK NRGKSDNVPS  420
EEVVKKMKNY WRQLLNAKLI TQRKFDNLTK AERGGLSGSE GYTSDEEVLE VFRNTLNKNS  480
EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK  540
KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS  600
EKLFDADFVL EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL  660
AYDILLKVDH IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK  720
YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI  780
QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE  840
VEEQGYKVSF ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN  900
HGQIRLSGGA ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED  960
QYELHIPIAI NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE  1020
QYSLNEIINN FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC  1080
ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK  1140
GYQITNKFES FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD  1200
RIMYVPEEDL FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT  1260
SAYKELFNKY GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS  1320
PVKNSDGIFY DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA  1380
ISNKEWLEYA QTSVKHGGGG SGGGGSGGGG SGPKKKRKVA AAGSEELLSK NYHLENEVAR  1440
LKKGSGSGGS GSGGSGSGSG GSGSGGSGSG EELLSKNYHL ENEVARLKKG SGSGGSGSGG  1500
SGSGSGGSGS GGSGSGEELL SKNYHLENEV ARLKKGSGSG GSGSGGSGSG SGGSGSGGSG  1560
SGEELLSKNY HLENEVARLK KGSGSGGSGS GGSGSGSGGS GSGGSGSGEE LLSKNYHLEN  1620
EVARLKKGSG SGGSGSGGSG SGSGGSGSGG SGSGEELLSK NYHLENEVAR LKKGSGSGGS  1680
GSGGSGSGSG GSGSGGSGSG EELLSKNYHL ENEVARLKKG SGSGGSGSGG SGSGSGGSGS  1740
GGSGSGEELL SKNYHLENEV ARLKKSGGSK RTADGSEFEP KKKRKVGSG              1789

SEQ ID NO: 141         moltype = AA  length = 1739
FEATURE                Location/Qualifiers
REGION                 1..1739
                       note = Synthetic polypeptide
source                 1..1739
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 141
MKRTADGSEF ESPKKKRKVE ASPASGPRHL MDPHIFTSNF NNGIGRHKTY LCYEVERLDN  60
GTSVKMDQHR GFLHNQAKNL LCGFYGRHAE LRFLDLVPSL QLDPAQIYRV TWFISWSPCF  120
SWGCAGEVRA FLQENTHVRL RIFAARIYDY DPLYKEALQM LRDAGAQVSI MTYDEFKHCW  180
DTFVDHQGCP FQPWDGLDEH SQALSGRLRA ILQNQGNSTS QSDGSSVPAD IDQSSDSDQS  240
SSQGQPGSKL EKFTNCYSLS KTLRFKAIPV GKTQENIDNK RLLVEDEKRA EDYKGVKKLL  300
DRYYLSFIND VLHSIKLKNL NNYISLFRKK TRTEKENKEL ENLEINLRKE IAKAFKGNEG  360
YKSLFKKDII ETILPEFLDD KDEIALVNSF NGFTTAFTGF FDNRENMFSE EAKSTSIAFR  420
CINENLTRYI SNMDIFEKVD AIFDKHEVQE IKEKILNSDY DVEDFFEGEF FNFVLTQEGI  480
DVYNAIIGGF VTESGEKIKG LNEYINLYNQ KTKQKLPKFK PLYKQVLSDR ESLSFYGEGS  540
SGENQTTQKG QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY  600
VDQELDINRL SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW  660
RQLLNAKLIT QRKFDNLTKA ERGGLSGYTS DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK  720
NFDEYSSAGI FVKNGPAIST ISKDIFGEWN VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR  780
KSFKKIGSFS LEQLQEYADA DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL  840
KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA  900
IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL  960
QKIDKDDVNG NYEKINYKLL PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM  1020
FNLNDCHKLI DFFKDSISRY PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS  1080
KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM  1140
RRASLKKEEL VVHPANSPIA NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP  1200
KNIFKINTEV RVLLKHDDNP YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI  1260
RIKTDYHSLL DKKEKERFEA RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED  1320
LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM  1380
STQNGFIFYI PAWLTSKIDP STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA  1440
LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY  1500
QQGDIRALLC EQSDKAFYSS FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN  1560
YEAQENAILP KNADANGAYN IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV  1620
KHSGGSGGSG GSTNLSDIIE KETGKQLVIQ ESILMLPEEV EEVIGNKPES DILVHTAYDE  1680
STDENVMLLT SDAPEYKPWA LVIQDSNGEN KIKMLSGGSK RTADGSEFEP KKKRKVGSG   1739

SEQ ID NO: 142         moltype = AA  length = 1739
FEATURE                Location/Qualifiers
REGION                 1..1739
                       note = Synthetic polypeptide
source                 1..1739
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 142
MKRTADGSEF ESPKKKRKVE ASPASGPRHL MDPHIFTSNF NNGIGRHKTY LCYEVERLDN  60
GTSVKMDQHR GFLHNQAKNL LCGFYGRHAE LRFLDLVPSL QLDPAQIYRV TWFISWSPCF  120
SWGCAGEVRA FLQENTHVRL RIFAARIYDY DPLYKEALQM LRDAGAQVSI MTYDEFKHCW  180
DTFVDHQGCP FQPWDGLDEH SQALSGRLRA ILQNQGNSTS QSDGSSVPAD IDQSSDSDQS  240
```

-continued

```
SSQGQPGSKL EKFTNCYSLS KTLRFKAIPV GKTQENIDNK RLLVEDEKRA EDYKGVKKLL  300
DRYYLSFIND VLHSIKLKNL NNYISLFRKK TRTEKENKEL ENLEINLRKE IAKAFKGNEG  360
YKSLFKKDII ETILPEFLDD KDEIALVNSF NGFTTAFTGF FDNRENMFSE EAKSTSIAFR  420
CINENLTRYI SNMDIFEKVD AIFDKHEVQE IKEKILNSDY DVEDFFEGEF FNFVLTQEGI  480
DVYNAIIGGF VTESGEKIKG LNEYINLYNQ KTKQKLPKFK PLYKQVLSDR ESLSFYGGSS  540
GENQTTQKGQ KNSRERMKRI EEGIKELGSQ ILKEHPVENT QLQNEKLYLY YLQNGRDMYV  600
DQELDINRLS DYDVDHIVPQ SFLKDDSIDN KVLTRSDKNR GKSDNVPSEE VVKKMKNYWR  660
QLLNAKLITQ RKFDNLTKAE RGGLSEGYTS DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK  720
NFDEYSSAGI FVKNGPAIST ISKDIFGEWN VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR  780
KSFKKIGSFS LEQLQEYADA DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL  840
KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA  900
IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL  960
QKIDKDDVNG NYEKINYKLL PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM  1020
FNLNDCHKLI DFFKDSISRY PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS  1080
KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM  1140
RRASLKKEEL VVHPANSPIA NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP  1200
KNIFKINTEV RVLLKHDDNP YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI  1260
RIKTDYHSLL DKKEKERFEA RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED  1320
LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM  1380
STQNGFIFYI PAWLTSKIDP STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA  1440
LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY  1500
QQGDIRALLC EQSDKAFYSS FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN  1560
YEAQENAILP KNADANGAYN IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV  1620
KHSGGSGGSG GSTNLSDIIE KETGKQLVIQ ESILMLPEEV EEVIGNKPES DILVHTAYDE  1680
STDENVMLLT SDAPEYKPWA LVIQDSNGEN KIKMLSGGSK RTADGSEFEP KKKRKVGSG   1739

SEQ ID NO: 143          moltype = AA  length = 1741
FEATURE                 Location/Qualifiers
REGION                  1..1741
                        note = Synthetic polypeptide
source                  1..1741
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 143
MKRTADGSEF ESPKKKRKVE ASPASGPRHL MDPHIFTSNF NNGIGRHKTY LCYEVERLDN  60
GTSVKMDQHR GFLHNQAKNL LCGFYGRHAE LRFLDLVPSL QLDPAQIYRV TWFISWSPCF  120
SWGCAGEVRA FLQENTHVRL RIFAARIYDY DPLYKEALQM LRDAGAQVSI MTYDEFKHCW  180
DTFVDHQGCP FQPWDGLDEH SQALSGRLRA ILQNQGNSTS QSDGSSVPAD IDQSSDSDQS  240
SSQGQPGSKL EKFTNCYSLS KTLRFKAIPV GKTQENIDNK RLLVEDEKRA EDYKGVKKLL  300
DRYYLSFIND VLHSIKLKNL NNYISLFRKK TRTEKENKEL ENLEINLRKE IAKAFKGNEG  360
YKSLFKKDII ETILPEFLDD KDEIALVNSF NGFTTAFTGF FDNRENMFSE EAKSTSIAFR  420
CINENLTRYI SNMDIFEKVD AIFDKHEVQE IKEKILNSDY DVEDFFEGEF FNFVLTQEGI  480
DVYNAIIGGF VTESGEKIKG LNEYINLYNQ KTKQKLPKFK PLYKQVLSDR ESLSFYGGSS  540
GENQTTQKGQ KNSRERMKRI EEGIKELGSQ ILKEHPVENT QLQNEKLYLY YLQNGRDMYV  600
DQELDINRLS DYDVDHIVPQ SFLKDDSIDN KVLTRSDKNR GKSDNVPSEE VVKKMKNYWR  660
QLLNAKLITQ RKFDNLTKAE RGGLSGSEGY TSDEEVLEVF RNTLNKNSEI FSSIKKLEKL  720
FKNFDEYSSA GIFVKNGPAI STISKDIFGE WNVIRDKWNA EYDDIHLKKK AVVTEKYEDD  780
RRKSFKKIGS FSLEQLQEYA DADLSVVEKL KEIIIQKVDE IYKVYGSSEK LFDADFVLEK  840
SLKKNDAVVA IMKDLLDSVK SFENYIKAFF GEGKETNRDE SFYGDFVLAY DILLKVDHIY  900
DAIRNYVTQK PYSKDKFKLY FQNPQFMGGW DKDKETDYRA TILRYGSKYY LAIMDKKYAK  960
CLQKIDKDDV NGNYEKINYK LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK IYKNGTFKKG  1020
DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD FNFSETEKYK DIAGFYREVE EQGYKVSFES  1080
ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH GTPNLHTMYF KLLFDENNHG QIRLSGGAEL  1140
FMRRASLKKE ELVVHPANSP IANKNPDNPK KTTTLSYDVY KDKRFSEDQY ELHIPIAINK  1200
CPKNIFKINT EVRVLLKHDD NPYVIGIARG ERNLLYIVVV DGKGNIVEQY SLNEIINNFN  1260
GIRIKTDYHS LLDKKEKERF EARQNWTSIE NIKELKAGYI SQVVHKICEL VEKYDAVIAL  1320
EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL NYMVDKKSNP CATGGALKGY QITNKFESFK  1380
SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL KTKYTSIADS KKFISSFDRI MYVPEEDLFE  1440
FALDYKNFSR TDADYIKKWK LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA YKELFNKYGI  1500
NYQQGDIRAL LCEQSDKAFY SSFMALMSLM LQMRNSITGR TDVDFLISPV KNSDGIFYDS  1560
RNYEAQENAI LPKNADANGA YNIARKVLWA IGQFKKAEDE KLDKVKIAIS NKEWLEYAQT  1620
SVKHSGGSGG SGGSTNLSDI IEKETGKQLV IQESILMLPE EVEEVIGNKP ESDILVHTAY  1680
DESTDENVML LTSDAPEYKP WALVIQDSNG ENKIKMLSGG SKRTADGSEF EPKKKRKVGS  1740
G                                                                 1741

SEQ ID NO: 144          moltype = AA  length = 1734
FEATURE                 Location/Qualifiers
REGION                  1..1734
                        note = Synthetic polypeptide
source                  1..1734
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 144
MKRTADGSEF ESPKKKRKVM EASPASGPRH LMDPHIFTSN FNNGIGRHKT YLCYEVERLD  60
NGTSVKMDQH RGFLHNQAKN LLCGFYGRHA ELRFLDLVPS LQLDPAQIYR VTWFISWSPC  120
FSWGCAGEVR AFLQENTHVR LRIFAARIYD YDPLYKEALQ MLRDAGAQVS IMTYDEFKHC  180
```

-continued

```
WDTFVDHQGC PFQPWDGLDE HSQALSGRLR AILQNQGNSG SETPGTSESA TPESMSKLEK  240
FTNCYSLSKT LRFKAIPVGK TQENIDNKRL LVEDEKRAED YKGVKKLLDR YYLSFINDVL  300
HSIKLKNLNN YISLFRKKTR TEKENKELEN LEINLRKEIA KAFKGNEGYK SLFKKDIIET  360
ILPEFLDDKD EIALVNSFNG FTTAFTGFFD NRENMFSEEA KSTSIAFRCI NENLTRYISN  420
MDIFEKVDAI FDKHEVQEIK EKILNSDYDV EDFFEGEFFN FVLTQEGIDV YNAIIGGFVT  480
ESGEKIKGLN EYINLYNQKT KQKLPKFKPL YKQVLSDRES LSFYGEGSSG ENQTTQKGQK  540
NSRERMKRIE EGIKELGSQI LKEHPVENTQ LQNEKLYLYY LQNGRDMYVD QELDINRLSD  600
YDVDHIVPQS FLKDDSIDNK VLTRSDKNRG KSDNVPSEEV VKKMKNYWRQ LLNAKLITQR  660
KFDNLTKAER GGLSGYTSDE EVLEVFRNTL NKNSEIFSSI KKLEKLFKNF DEYSSAGIFV  720
KNGPAISTIS KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT EKYEDDRRKS FKKIGSFSLE  780
QLQEYADADL SVVEKLKEII IQKVDEIYKV YGSSEKLFDA DFVLEKSLKK NDAVVAIMKD  840
LLDSVKSFEN YIKAFFGEGK ETNRDESFYG DFVLAYDILL KVDHIYDAIR NYVTQKPYSK  900
DKFKLYFQNP QFMGGWDKDK ETDYRATILR YGSKYYLAIM DKKYAKCLQK IDKDDVNGNY  960
EKINYKLLPG PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFN LNDCHKLIDF  1020
FKDSISRYPK WSNAYDFNFS ETEKYKDIAG FYREVEEQGY KVSFESASKK EVDKLVEEGK  1080
LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF DENNHGQIRL SGGAELFMRR ASLKKEELVV  1140
HPANSPIANK NPDNPKKTTT LSYDVYKDKR FSEDQYELHI PIAINKCPKN IFKINTEVRV  1200
LLKHDDNPYV IGIARGERNL LYIVVVDGKG NIVEQYSLNE IINNFNGIRI KTDYHSLLDK  1260
KEKERFEARQ NWTSIENIKE LKAGYISQVV HKICELVEKY DAVIALEDLN SGFKNSRVKV  1320
EKQVYQKFEK MLIDKLNYMV DKKSNPCATG GALKGYQITN KFESFKSMST QNGFIFYIPA  1380
WLTSKIDPST GFVNLLKTKY TSIADSKKFI SSFDRIMYVP EEDLFEFALD YKNFSRTDAD  1440
YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL FNKYGINYQQ GDIRALLCEQ  1500
SDKAFYSSFM ALMSLMLQMR NSITGRTDVD FLISPVKNSD GIFYDSRNYE AQENAILPKN  1560
ADANGAYNIA RKVLWAIGQF KKAEDEKLDK VKIAISNKEW LEYAQTSVKH GSPKKKRKVS  1620
GGSTNLSDII EKETGKQLVI QESILMLPEE VEEVIGNKPE SDILVHTAYD ESTDENVMLL  1680
TSDAPEYKPW ALVIQDSNGE NKIKMLTKYD SGGSKRTADG SEFEPKKKRK VGSG        1734

SEQ ID NO: 145          moltype = AA  length = 1734
FEATURE                 Location/Qualifiers
REGION                  1..1734
                        note = Synthetic polypeptide
source                  1..1734
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 145
MKRTADGSEF ESPKKKRKVM EASPASGPRH LMDPHIFTSN FNNGIGRHKT YLCYEVERLD  60
NGTSVKMDQH RGFLHNQAKN LLCGFYGRHA ELRFLDLVPS LQLDPAQIYR VTWFISWSPC  120
FSWGCAGEVR AFLQENTHVR LRIFAARIYD YDPLYKEALQ MLRDAGAQVS IMTYDEFKHC  180
WDTFVDHQGC PFQPWDGLDE HSQALSGRLR AILQNQGNSG SETPGTSESA TPESMSKLEK  240
FTNCYSLSKT LRFKAIPVGK TQENIDNKRL LVEDEKRAED YKGVKKLLDR YYLSFINDVL  300
HSIKLKNLNN YISLFRKKTR TEKENKELEN LEINLRKEIA KAFKGNEGYK SLFKKDIIET  360
ILPEFLDDKD EIALVNSFNG FTTAFTGFFD NRENMFSEEA KSTSIAFRCI NENLTRYISN  420
MDIFEKVDAI FDKHEVQEIK EKILNSDYDV EDFFEGEFFN FVLTQEGIDV YNAIIGGFVT  480
ESGEKIKGLN EYINLYNQKT KQKLPKFKPL YKQVLSDRES LSFYGGSSGE NQTTQKGQKN  540
SRERMKRIEE GIKELGSQIL KEHPVENTQL QNEKLYLYYL QNGRDMYVDQ ELDINRLSDY  600
DVDHIVPQSF LKDDSIDNKV LTRSDKNRGK SDNVPSEEVV KKMKNYWRQL LNAKLITQRK  660
FDNLTKAERG GLSEGYTSDE EVLEVFRNTL NKNSEIFSSI KKLEKLFKNF DEYSSAGIFV  720
KNGPAISTIS KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT EKYEDDRRKS FKKIGSFSLE  780
QLQEYADADL SVVEKLKEII IQKVDEIYKV YGSSEKLFDA DFVLEKSLKK NDAVVAIMKD  840
LLDSVKSFEN YIKAFFGEGK ETNRDESFYG DFVLAYDILL KVDHIYDAIR NYVTQKPYSK  900
DKFKLYFQNP QFMGGWDKDK ETDYRATILR YGSKYYLAIM DKKYAKCLQK IDKDDVNGNY  960
EKINYKLLPG PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFN LNDCHKLIDF  1020
FKDSISRYPK WSNAYDFNFS ETEKYKDIAG FYREVEEQGY KVSFESASKK EVDKLVEEGK  1080
LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF DENNHGQIRL SGGAELFMRR ASLKKEELVV  1140
HPANSPIANK NPDNPKKTTT LSYDVYKDKR FSEDQYELHI PIAINKCPKN IFKINTEVRV  1200
LLKHDDNPYV IGIARGERNL LYIVVVDGKG NIVEQYSLNE IINNFNGIRI KTDYHSLLDK  1260
KEKERFEARQ NWTSIENIKE LKAGYISQVV HKICELVEKY DAVIALEDLN SGFKNSRVKV  1320
EKQVYQKFEK MLIDKLNYMV DKKSNPCATG GALKGYQITN KFESFKSMST QNGFIFYIPA  1380
WLTSKIDPST GFVNLLKTKY TSIADSKKFI SSFDRIMYVP EEDLFEFALD YKNFSRTDAD  1440
YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL FNKYGINYQQ GDIRALLCEQ  1500
SDKAFYSSFM ALMSLMLQMR NSITGRTDVD FLISPVKNSD GIFYDSRNYE AQENAILPKN  1560
ADANGAYNIA RKVLWAIGQF KKAEDEKLDK VKIAISNKEW LEYAQTSVKH GSPKKKRKVS  1620
GGSTNLSDII EKETGKQLVI QESILMLPEE VEEVIGNKPE SDILVHTAYD ESTDENVMLL  1680
TSDAPEYKPW ALVIQDSNGE NKIKMLTKYD SGGSKRTADG SEFEPKKKRK VGSG        1734

SEQ ID NO: 146          moltype = AA  length = 1736
FEATURE                 Location/Qualifiers
REGION                  1..1736
                        note = Synthetic polypeptide
source                  1..1736
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 146
MKRTADGSEF ESPKKKRKVM EASPASGPRH LMDPHIFTSN FNNGIGRHKT YLCYEVERLD  60
NGTSVKMDQH RGFLHNQAKN LLCGFYGRHA ELRFLDLVPS LQLDPAQIYR VTWFISWSPC  120
FSWGCAGEVR AFLQENTHVR LRIFAARIYD YDPLYKEALQ MLRDAGAQVS IMTYDEFKHC  180
```

-continued

```
WDTFVDHQGC PFQPWDGLDE HSQALSGRLR AILQNQGNSG SETPGTSESA TPESMSKLEK  240
FTNCYSLSKT LRFKAIPVGK TQENIDNKRL LVEDEKRAED YKGVKKLLDR YYLSFINDVL  300
HSIKLKNLNN YISLFRKKTR TEKENKELEN LEINLRKEIA KAFKGNEGYK SLFKKDIIET  360
ILPEFLDDKD EIALVNSFNG FTTAFTGFFD NRENMFSEEA KSTSIAFRCI NENLTRYISN  420
MDIFEKVDAI FDKHEVQEIK EKILNSDYDV EDFFEGEFFN FVLTQEGIDV YNAIIGGFVT  480
ESGEKIKGLN EYINLYNQKT KQKLPKFKPL YKQVLSDRES LSFYGGSSGE NQTTQKGQKN  540
SRERMKRIEE GIKELGSQIL KEHPVENTQL QNEKLYLYYL QNGRDMYVDQ ELDINRLSDY  600
DVDHIVPQSF LKDDSIDNKV LTRSDKNRGK SDNVPSEEVV KKMKNYWRQL LNAKLITQRK  660
FDNLTKAERG GLSGSEGYTS DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI  720
FVKNGPAIST ISKDIFGEWN VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS  780
LEQLQEYADA DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM  840
KDLLDSVKSF ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY  900
SKDKFKLYFQ NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG  960
NYEKINYKLL PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI  1020
DFFKDSISRY PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE  1080
GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL  1140
VVHPANSPIA NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV  1200
RVLLKHDDNP YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL  1260
DKKEKERFEA RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV  1320
KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI  1380
PAWLTSKIDP STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD  1440
ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC  1500
EQSDKAFYSS FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP  1560
KNADANGAYN IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KHGSPKKKRK  1620
VSGGSTNLSD IIEKETGKQL VIQESILMLP EEVEEVIGNK PESDILVHTA YDESTDENVM  1680
LLTSDAPEYK PWALVIQDSN GENKIKMLTK YDSGGSKRTA DGSEFEPKKK RKVGSG      1736

SEQ ID NO: 147          moltype = AA  length = 1616
FEATURE                 Location/Qualifiers
REGION                  1..1616
                        note = Synthetic polypeptide
source                  1..1616
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 147
MKRTADGSEF ESPKKKRKVS EVEFSHEYWM RHALTLAKRA RDEREVPVGA VLVLNNRVIG  60
EGWNRAIGLH DPTAHAEIMA LRQGGLVMQN YRLIDATLYV TFEPCVMCAG AMIHSRIGRV  120
VFGVRNSKRG AAGSLMNVLN YPGMNHRVEI TEGILADECA ALLCDFYRMP RQVFNAQKKA  180
QSSINSGGSS GGSSGSETPG TSESATPESS GGSSGGSSKL EKFTNCYSLS KTLRFKAIPV  240
GKTQENIDNK RLLVEDEKRA EDYKGVKKLL DRYYLSFIND VLHSIKLKNL NNYISLFRKK  300
TRTEKENKEL ENLEINLRKE IAKAFKGNEG YKSLFKKDII ETILPEFLDD KDEIALVNSF  360
NGFTTAFTGF FDNRENMFSE EAKSTSIAFR CINENLTRYI SNMDIFEKVD AIFDKHEVQE  420
IKEKILNSDY DVEDFFEGEF FNFVLTQEGI DVYNAIIGGF VTESGEKIKG LNEYINLYNQ  480
KTKQKLPKFK PLYKQVLSDR ESLSFYGEGS SGENQTTQKG QKNSRERMKR IEEGIKELGS  540
QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL SDYDVDHIVP QSFLKDDSID  600
NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT QRKFDNLTKA ERGGLSGYTS  660
DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI FVKNGPAIST ISKDIFGEWN  720
VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS LEQLQEYADA DLSVVEKLKE  780
IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE  840
GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK  900
DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL PGPNKMLPKV  960
FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY PKWSNAYDFN  1020
FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT  1080
PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA NKNPDNPKKT  1140
TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP YVIGIARGER  1200
NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA RQNWTSIENI  1260
KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY  1320
MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP STGFVNLLKT  1380
KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR  1440
NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS FMALMSLMLQ  1500
MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN IARKVLWAIG  1560
QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KHSGGSKRTA DGSEFEPKKK RKVGSG      1616

SEQ ID NO: 148          moltype = AA  length = 1616
FEATURE                 Location/Qualifiers
REGION                  1..1616
                        note = Synthetic polypeptide
source                  1..1616
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 148
MKRTADGSEF ESPKKKRKVS EVEFSHEYWM RHALTLAKRA RDEREVPVGA VLVLNNRVIG  60
EGWNRAIGLH DPTAHAEIMA LRQGGLVMQN YRLIDATLYV TFEPCVMCAG AMIHSRIGRV  120
VFGVRNSKRG AAGSLMNVLN YPGMNHRVEI TEGILADECA ALLCDFYRMP RQVFNAQKKA  180
QSSINSGGSS GGSSGSETPG TSESATPESS GGSSGGSSKL EKFTNCYSLS KTLRFKAIPV  240
GKTQENIDNK RLLVEDEKRA EDYKGVKKLL DRYYLSFIND VLHSIKLKNL NNYISLFRKK  300
```

-continued

```
TRTEKENKEL ENLEINLRKE IAKAFKGNEG YKSLFKKDII ETILPEFLDD KDEIALVNSF  360
NGFTTAFTGF FDNRENMFSE EAKSTSIAFR CINENLTRYI SNMDIFEKVD AIFDKHEVQE  420
IKEKILNSDY DVEDFFEGEF FNFVLTQEGI DVYNAIIGGF VTESGEKIKG LNEYINLYNQ  480
KTKQKLPKFK PLYKQVLSDR ESLSFYGGSS GENQTTQKGQ KNSRERMKRI EEGIKELGSQ  540
ILKEHPVENT QLQNEKLYLY YLQNGRDMYV DQELDINRLS DYDVDHIVPQ SFLKDDSIDN  600
KVLTRSDKNR GKSDNVPSEE VVKKMKNYWR QLLNAKLITQ RKFDNLTKAE RGGLSEGYTS  660
DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI FVKNGPAIST ISKDIFGEWN  720
VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS LEQLQEYADA DLSVVEKLKE  780
IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE  840
GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK  900
DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL PGPNKMLPKV  960
FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY PKWSNAYDFN  1020
FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT  1080
PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA NKNPDNPKKT  1140
TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP YVIGIARGER  1200
NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA RQNWTSIENI  1260
KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY  1320
MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP STGFVNLLKT  1380
KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR  1440
NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS FMALMSLMLQ  1500
MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN IARKVLWAIG  1560
QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KHSGGSKRTA DGSEFEPKKK RKVGSG      1616

SEQ ID NO: 149         moltype = AA  length = 1618
FEATURE                Location/Qualifiers
REGION                 1..1618
                       note = Synthetic polypeptide
source                 1..1618
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 149
MKRTADGSEF ESPKKKRKVS EVEFSHEYWM RHALTLAKRA RDEREVPVGA VLVLNNRVIG  60
EGWNRAIGLH DPTAHAEIMA LRQGGLVMQN YRLIDATLYV TFEPCVMCAG AMIHSRIGRV  120
VFGVRNSKRG AAGSLMNVLN YPGMNHRVEI TEGILADECA ALLCDFYRMP RQVFNAQKKA  180
QSSINSGGSS GGSSGSETPG TSESATPESS GGSSGGSSKL EKFTNCYSLS KTLRFKAIPV  240
GKTQENIDNK RLLVEDEKRA EDYKGVKKLL DRYYLSFIND VLHSIKLKNL NNYISLFRKK  300
TRTEKENKEL ENLEINLRKE IAKAFKGNEG YKSLFKKDII ETILPEFLDD KDEIALVNSF  360
NGFTTAFTGF FDNRENMFSE EAKSTSIAFR CINENLTRYI SNMDIFEKVD AIFDKHEVQE  420
IKEKILNSDY DVEDFFEGEF FNFVLTQEGI DVYNAIIGGF VTESGEKIKG LNEYINLYNQ  480
KTKQKLPKFK PLYKQVLSDR ESLSFYGGSS GENQTTQKGQ KNSRERMKRI EEGIKELGSQ  540
ILKEHPVENT QLQNEKLYLY YLQNGRDMYV DQELDINRLS DYDVDHIVPQ SFLKDDSIDN  600
KVLTRSDKNR GKSDNVPSEE VVKKMKNYWR QLLNAKLITQ RKFDNLTKAE RGGLSGSEGY  660
TSDEEVLEVF RNTLNKNSEI FSSIKKLEKL FKNFDEYSSA GIFVKNGPAI STISKDIFGE  720
WNVIRDKWNA EYDDIHLKKK AVVTEKYEDD RRKSFKKIGS FSLEQLQEYA DADLSVVEKL  780
KEIIIQKVDE IYKVYGSSEK LFDADFVLEK SLKKNDAVVA IMKDLLDSVK SFENYIKAFF  840
GEGKETNRDE SFYGDFVLAY DILLKVDHIY DAIRNYVTQK PYSKDKFKLY FQNPQFMGGW  900
DKDKETDYRA TILRYGSKYY LAIMDKKYAK CLQKIDKDDV NGNYEKINYK LLPGPNKMLP  960
KVFFSKKWMA YYNPSEDIQK IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD  1020
FNFSETEKYK DIAGFYREVE EQGYKVSFES ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH  1080
GTPNLHTMYF KLLFDENNHG QIRLSGGAEL FMRRASLKKE ELVVHPANSP IANKNPDNPK  1140
KTTTLSYDVY KDKRFSEDQY ELHIPIAINK CPKNIFKINT EVRVLLKHDD NPYVIGIARG  1200
ERNLLYIVVV DGKGNIVEQY SLNEIINNFN GIRIKTDYHS LLDKKEKERF EARQNWTSIE  1260
NIKELKAGYI SQVVHKICEL VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL  1320
NYMVDKKSNP CATGGALKGY QITNKFESFK SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL  1380
KTKYTSIADS KKFISSFDRI MYVPEEDLFE FALDYKNFSR TDADYIKKWK LYSYGNRIRI  1440
FRNPKKNNVF DWEEVCLTSA YKELFNKYGI NYQQGDIRAL LCEQSDKAFY SSFMALMSLM  1500
LQMRNSITGR TDVDFLISPV KNSDGIFYDS RNYEAQENAI LPKNADANGA YNIARKVLWA  1560
IGQFKKAEDE KLDKVKIAIS NKEWLEYAQT SVKHSGGSKR TADGSEFEPK KKRKVGSG    1618

SEQ ID NO: 150         moltype = AA  length = 173
FEATURE                Location/Qualifiers
source                 1..173
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 150
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFR         173

SEQ ID NO: 151         moltype = AA  length = 1053
FEATURE                Location/Qualifiers
source                 1..1053
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
```

-continued

```
SEQUENCE: 151
INENLTRYIS NMDIFEKVDA IFDKHEVQEI KEKILNSDYD VEDFFEGEFF NFVLTQEGID  60
VYNAIIGGFV TESGEKIKGL NEYINLYNQK TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT  120
SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG IFVKNGPAIS TISKDIFGEW  180
NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK  240
EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI MKDLLDSVKS FENYIKAFFG  300
EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD  360
KDKETDYRAT ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN GNYEKINYKL LPGPNKMLPK  420
VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL IDFFKDSISR YPKWSNAYDF  480
NFSETEKYKD IAGFYREVEE QGYKVSFESA SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG  540
TPNLHTMYFK LLFDENNHGQ IRLSGGAELF MRRASLKKEE LVVHPANSPI ANKNPDNPKK  600
TTTLSYDVYK DKRFSEDQYE LHIPIAINKC PKNIFKINTE VRVLLKHDDN PYVIGIDRGE  660
RNLLYIVVVD GKGNIVEQYS LNEIINNFNG IRIKTDYHSL LDKKEKERFE ARQNWTSIEN  720
IKELKAGYIS QVVHKICELV EKYDAVIALE DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN  780
YMVDKKSNPC ATGGALKGYQ ITNKFESFKS MSTQNGFIFY IPAWLTSKID PSTGFVNLLK  840
TKYTSIADSK KFISSFDRIM YVPEEDLFEF ALDYKNFSRT DADYIKKWKL YSYGNRIRIF  900
RNPKKNNVFD WEEVCLTSAY KELFNKYGIN YQQGDIRALL CEQSDKAFYS SFMALMSLML  960
QMRNSITGRT DVDFLISPVK NSDGIFYDSR NYEAQENAIL PKNADANGAY NIARKVLWAI  1020
GQFKKAEDEK LDKVKIAISN KEWLEYAQTS VKH                              1053

SEQ ID NO: 152         moltype = AA  length = 310
FEATURE                Location/Qualifiers
source                 1..310
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 152
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGYTSDEEVL  300
EVFRNTLNKC                                                        310

SEQ ID NO: 153         moltype = AA  length = 918
FEATURE                Location/Qualifiers
source                 1..918
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 153
NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH  60
LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG  120
SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF  180
VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG  240
SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE  300
DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY  360
REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE  420
NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS  480
EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI  540
VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK  600
ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA  660
LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS  720
FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC  780
LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL  840
ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK  900
IAISNKEWLE YAQTSVKH                                               918

SEQ ID NO: 154         moltype = AA  length = 406
FEATURE                Location/Qualifiers
source                 1..406
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 154
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGYTSDEEVL  300
EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK  360
WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ EYADAC                406

SEQ ID NO: 155         moltype = AA  length = 822
FEATURE                Location/Qualifiers
source                 1..822
```

-continued

```
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 155
DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF  60
ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ  120
NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL  180
PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY  240
PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN  300
KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA  360
NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP  420
YVIGIDRGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA  480
RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF  540
EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP  600
STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY  660
SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS  720
FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN  780
IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KH                     822

SEQ ID NO: 156         moltype = AA  length = 441
FEATURE                Location/Qualifiers
source                 1..441
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 156
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGYTSDEEVL  300
EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK  360
WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ EYADADLSVV EKLKEIIIQK  420
VDEIYKVYGS SEKLFDADFV C                                            441

SEQ ID NO: 157         moltype = AA  length = 787
FEATURE                Location/Qualifiers
source                 1..787
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 157
LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD  60
HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK  120
YAKCLQKIDK DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF  180
KKGDMFNLND CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS  240
FESASKKEVD KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG  300
AELFMRRASL KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA  360
INKCPKNIFK INTEVRVLLK HDDNPYVIGI DRGERNLLYI VVVDGKGNIV EQYSLNEIIN  420
NFNGIRIKTD YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV  480
IALEDLNSGF KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE  540
SFKSMSTQNG FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED  600
LFEFALDYKN FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK  660
YGINYQQGDI RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF  720
YDSRNYEAQE NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY  780
AQTSVKH                                                            787

SEQ ID NO: 158         moltype = AA  length = 550
FEATURE                Location/Qualifiers
source                 1..550
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 158
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGYTSDEEVL  300
EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK  360
WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ EYADADLSVV EKLKEIIIQK  420
VDEIYKVYGS SEKLFDADFV LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN  480
RDESFYGDFV LAYDILLKVD HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD  540
YRATILRYGC                                                         550

SEQ ID NO: 159         moltype = AA  length = 678
```

-continued

```
FEATURE                Location/Qualifiers
source                 1..678
                       mol_type = protein
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 159
SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE  60
DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY  120
REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE  180
NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS  240
EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY IVVVDGKGNI  300
VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK  360
ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA  420
LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS  480
FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC  540
LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL  600
ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK  660
IAISNKEWLE YAQTSVKH                                                678

SEQ ID NO: 160         moltype = AA  length = 683
FEATURE                Location/Qualifiers
source                 1..683
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 160
TLNIEDEHRL HETSKEPDVS LGSTWLSDFP QAWAETGGMG LAVRQAPLII PLKATSTPVS  60
IKQYPMSQEA RLGIKPHIQR LLDQGILVPC QSPWNTPLLP VKKPGTNDYR PVQDLREVNK  120
RVEDIHPTVP NPYNLLSGLP PSHQWYTVLD LKDAFFCLRL HPTSQPLFAF EWRDPEMGIS  180
GQLTWTRLPQ GFKNSPTLFN EALHRDLADF RIQHPDLILL QYVDDLLLAA TSELDCQQGT  240
RALLQTLGNL GYRASAKKAQ ICQKQVKYLG YLLKEGQRWL TEARKETVMG QPTPKTPRQL  300
REFLGKAGFC RLFIPGFAEM AAPLYPLTKP GTLFNWGPDQ QKAYQEIKQA LLTAPALGLP  360
DLTKPFELFV DEKQGYAKGV LTQKLGPWRR PVAYLSKKLD PVAAGWPPCL RMVAAIAVLT  420
KDAGKLTMGQ PLVILAPHAV EALVKQPPDR WLSNARMTHY QALLLDTDRV QFGPVVALNP  480
ATLLPLPEEG LQHNCLDILA EAHGTRPDLT DQPLPDADHT WYTDGSSLLQ EGQRKAGAAV  540
TTETEVIWAK ALPAGTSAQR AELIALTQAL KMAEGKKLNV YTDSRYAFAT AHIHGEIYRR  600
RGWLTSEGKE IKNKDEILAL LKALFLPKRL SIIHCPGHQK GHSAEARGNR MADQAARKAA  660
ITETPDTSTL LIENSSPNSR LIN                                          683

SEQ ID NO: 161         moltype = DNA  length = 2031
FEATURE                Location/Qualifiers
misc_feature           1..2031
                       note = MMLV_MO1
source                 1..2031
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 161
accctgaaca tcgaggacga gtatcggctt catgagacca gcaaagagcc ggacgtctcc  60
ctgggaagca cctggctgag cgacttcccg caagcctggg cggagaccgg tggtatgggg  120
ctcgcagtgc ggcaagcgcc gttgataatc ccgctaaagg ccacgagcac gcccgtgtct  180
atcaagcagt acccgatgag tcaagaggca cgcctcggta tcaagccgca tatccagcgc  240
ctcctggacc agggcatcct cgtgccctgc cagtctccct ggaatacgcc tctgctaccc  300
gtcaagaagc ctggcaccaa cgattacagg ccggtgcaag acctgcgtga ggtcaacaag  360
cgcgtggagg acatccaccc aacggtgccc aacccgtaca atctcctatc tggccttccg  420
ccctcgcacc agtggtacac ggtcttggac ctaaaggacg cattcttctg tctgaggctg  480
caccctacgt cccagccgct gttcgccttc gagtggcgcg acccggagat gggcatctct  540
ggccagctaa cttggacgcg attgccccag gggtttaaga actcgcccac actcttcaac  600
gaggcactcc accgtgacct ggccgacttt cgcatacagc accccgacct tatcctgttg  660
cagtacgtcg atgacctgct cctggcggcc acgtccgagc tggactgcca gcaaggcacc  720
cgcgccctac ttcaaaccct gggcaacctg ggttaccgtg cgtccgccaa gaaggcccaa  780
atctgccaaa agcaagtcaa gtacctcggc tacctcttga aggagggaca gcgctggctg  840
acggaggcga ggaaggagac ggtgatgggt cagcccacac ccaagacccc gaggcagcta  900
agggagttcc tggggaaggc gggcttctgc cgtctattca tccctggctt cgcggagatg  960
gcggccccgc tgtacccgct aacgaagccg ggcacgctgt tcaactgggg ccctgaccag  1020
cagaaggcgt accaggagat caagcaagcg ttgcttactg ccccagcact cggcctcccc  1080
gacctcacaa agccgttcga gctattcgtt gacgagaaac agggctacgc gaagggtgtg  1140
ctgactcaaa agctagggcc gtggcgacgc ccagtagcct acctgagcaa gaagctcgac  1200
cccgtggcgg cgggctggcc accatgcctc cggatggtcg cggctatcgc ggtgctcaca  1260
aaggatgcgg ggaagctcac gatggggcag cccctggtga tcctggcccc acacgcggtg  1320
gaggcgttag tgaagcaacc tcccgaccga tggctgagca acgcccgcat gacccactac  1380
caggcgctcc tcctcgacac cgaccgcgtg caattcggcc ctgtcgtggc actgaacccg  1440
gccacgctgc tcccactgcc cgaggagggg ctacagcaca actgcctcga tatactggcg  1500
gaagcccacg gcacccgccc cgacttgacg gaccagccgc ttcccgacgc ggaccatacg  1560
tggtacaccg acgggagttc cttactccaa gagggccagc gaaaggcggg cgctgcggtg  1620
accactgaga cggaagtaat ctgggcaaag gcgctgcctg cgggcacgtc tgcccagcgg  1680
gcggagctga tcgccctgac ccaggccctc aagatggccg agggcaagaa gctgaacgtt  1740
tacaccgaca gtcggtatgc cttcgcaact gcccacatcc acggcgaaat ctaccgtcgg  1800
```

-continued

```
cgcggctggc tgacgagcga gggcaaggag atcaagaaca aggacgagat cctcgccctg  1860
ctaaaggcac tcttcctgcc caagcgactg tccatcattc actgtccggg gcaccagaag  1920
ggccattccg ccgaggcgcg gggcaaccgc atggcggatc aggccgctcg gaaggcggcg  1980
atcaccgaga cgcccgatac gagcacgctc ctgattgaaa actcgtcgcc g           2031

SEQ ID NO: 162          moltype = DNA  length = 2031
FEATURE                 Location/Qualifiers
misc_feature            1..2031
                        note = MMLV_MO2
source                  1..2031
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 162
accctgaaca ttgaggacga gtatcgcctg cacgagacca gcaaggagcc cgacgtgagc  60
ctcggctcaa cgtggctcag tgacttccca caagcctggg ccgagactgg tggaatgggg  120
ctggccgtgc gccaagctcc cctcatcatt cctctcaagg ccacttcgac gcctgtctcc  180
atcaagcagt accccatgtc ccaggaggct aggctgggca tcaagccgca catccaacgc  240
ctgttagatc aaggaatact ggtcccctgc cagtcgccgt ggaatactcc gctactgcct  300
gtcaagaagc cgggcacgaa cgactaccgg cctgtccaag acctgcgcga ggtgaacaag  360
cgggtcgagg acattcaccc caccgttccc aacccttaca acctattgtc tggcctccca  420
ccgagccatc agtggtacac cgtcctggac ctcaaggatg cgttcttctg tctgcggcta  480
caccctacgt cgcaaccact cttcgccttc gagtggcggg atccggagat ggggatctcg  540
gggcagctca cgtggactcg gctgcctcag gggttcaaga actccccaac actctttaac  600
gaggcactgc atcgggatct ggccgacttc cgtatccagc acccagacct catcctccta  660
cagtacgtgg acgacttgct gctggccgcg accagcgagc tggactgcca gcaagggaca  720
cgcgcgctgc tccagacgct cgggaacctg ggataccgcg ccagcgctaa gaaggctcaa  780
atctgtcaga aacaagtgaa gtacctgggt tacctgctca aggagggtca gcgttggctt  840
accgaggccc gcaaggagac cgtgatgggg caacccacgc caaagacgcc ccgacagcta  900
cgcgagttcc tgggcaaggc tgggttctgt cggttgttca tccccggttt cgctgagatg  960
gccgcccctc tctacccgct gacgaaacct gggactctgt tcaactgggg ccctgaccag  1020
cagaaggcgt accaggagat caagcaagcg ctgcttaccg ccccggcgct cggattgccg  1080
gaccttacca agcccttcga gctgttcgtg gacgagaagc aaggttacgc gaagggcgtc  1140
ctgacacaaa agttggggcc ctggcgtagg ccggtcgcct acctcagcaa gaagttggac  1200
cccgtggcgg cgggctggcc gccgtgcctc cgtatggtgg cagccatcgc cgttctcacc  1260
aaagacgctg gcaagctcac gatggggcag cctctggtga tcctggcacc ccatgccgtc  1320
gaggcgctcg tgaagcagcc gcctgaccgc tggctgagta acgcacggat gacccattac  1380
caagccctat tgctagacac ggatcgcgtc caatttgggc ccgtggtggc tctgaaccct  1440
gccactctcc ttcccctccc tgaggagggc ttgcagcata actgcctgga catactggct  1500
gaggcccacg ggacaaggcc ggacctaacg gaccagcctc taccggacgc ggatcacaca  1560
tggtacaccg acggctcctc tctcctacaa gaggggcagc ggaaggcggg tgccgccgtc  1620
accacggaga cggaggtgat ctgggctaag gcactgcccg ccgggacttc ggcacagcga  1680
gccgagctaa tagccctcac acaagcgctg aaaatggccg agggcaagaa gctaaacgtc  1740
tatacggact cccgatacgc tttcgccacc gcccacattc atggcgaaat ctaccgccgc  1800
cgtggctggc tgacgtccga gggcaaggag atcaagaaca aggacgagat cctcgccctc  1860
ctgaaagccc tgttcctgcc gaaaaggctt tcgataatcc actgccccgg ccaccagaag  1920
gggcactccg ccgaggcacg cggcaaccgt atggccgacc aggccgcccg gaaggcggcg  1980
atcacggaaa ccccggacac atccacgctc ctcatcgaga acagcagccc c           2031

SEQ ID NO: 163          moltype = DNA  length = 2031
FEATURE                 Location/Qualifiers
misc_feature            1..2031
                        note = MMLV_MO3
source                  1..2031
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 163
accctcaaca tcgaggacga gtaccgtctg cacgagacgt cgaaggagcc ggatgtctca  60
ctcggctcca cgtggctcag cgacttccca caggcgtggg cggagaccgg tggcatggga  120
ctggcggtgc gacaagcgcc ccttatcatt cccctaaagg cgacgtcaac tcctgtttcc  180
attaagcagt accctatgtc ccaggaggcc cggctcggca tcaagccaca catccaacgg  240
ctattggacc agggtatctt ggtgccgtgc caatccccgt ggaacactcc ccttctaccc  300
gtcaagaagc caggcaccaa cgactaccgc cccgtgcaag acctgcgcga ggtcaacaag  360
cgagtggagg acatccatcc taccgtcccg aacccgtaca acctgctttc cggcctcccg  420
ccctcgcacc agtggtacac cgttctcgac ttgaaggacg cattcttctg tctgcgcctc  480
cacccaacga gccagcccct cttcgctttc gagtggcgcg acccggagat gggaatttcg  540
ggccagctta catggacccg cctcccacag gggttcaaga acagcccgac gctgttcaac  600
gaggccctgc accgcgactt ggcagacttc cgaatccagc atcccgacct aatcctcctg  660
caatacgttg acgacttatt gctggccgcg accagcgagc ttgactgcca gcaagggact  720
cgcgcgcttc tacagacgct cgggaacctg ggctaccgtg cctcagctaa gaaggcccag  780
atttgccaga agcaagttaa gtatctcggc tacctcctca aggagggaca gcggtggctg  840
acggaggccc gcaaggagac ggtcatgggc cagccaacac cgaaaacgcc cagacaactc  900
cgcgagttcc tcggcaaagc gggcttctgt cggctgttta tccccggctt cgccgagatg  960
gccgcgcccc tctacccact gacgaaaccc ggcaccctgt tcaactgggg cccggatcag  1020
cagaaggcgt atcaggagat caagcaagca ctcctgacag ccccggccct gggattgccc  1080
gaccttacga agcccttcga gttattcgtg gacgagaagc aaggctatgc gaagggtgtc  1140
ctcacgcaga agctggggcc ctggaggcgg cccgtcgcgt acctaagcaa gaagctcgac  1200
```

-continued

```
ccagtggcgg cgggttggcc gccctgcctc cgcatggtgg ccgcgattgc ggttctcaca  1260
aaggacgccg ggaagctcac gatggggcag ccacttgtca tcctcgctcc gcacgccgtg  1320
gaggcactgg tgaagcagcc gccggatcgc tggctgtcta atgctcgcat gacccactat  1380
caggcgctgc tcctagacac tgacagggtt cagttcggcc ccgttgtcgc gcttaacccc  1440
gctacactac tcccgctgcc ggaggagggt ttgcagcata actgcctcga catcctcgcc  1500
gaggcccacg gcacgcgacc cgacctaacg gaccagccgc tgccggacgc tgaccacact  1560
tggtacaccg acggcagctc cctcctgcaa gagggacagc ggaaggccgg tgccgccgtg  1620
acgacggaga cggaggtgat atgggctaag gccctgcccg ctggtacgtc cgcccagcga  1680
gccgagctga tcgccctgac gcaagccctc aagatggccg agggcaagaa gctaaacgtc  1740
tatacggaca gccgctacgc attcgccaca gcccacattc acggagagat ataccggagg  1800
cgcggctggc ttacgtccga aggcaaggag attaagaaca aagatgagat tctggcgctg  1860
ttgaaggccc tcttcctccc caagcggctt tccatcatac actgtccagg ccaccagaag  1920
ggccactcgg cggaggcgcg gggcaaccgg atggccgacc aggcggcgcg caaagccgcg  1980
atcacggaga ccccagacac ttccacgctc ctgatcgaga acagtagccc c           2031

SEQ ID NO: 164         moltype = DNA  length = 2031
FEATURE                Location/Qualifiers
misc_feature           1..2031
                       note = MMLV_DO1
source                 1..2031
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 164
actttgaata tcgaagatga gtaccggcta catgagacgt ctaaggagcc tgatgtttca  60
ctcgggagca cttggctaag cgatttccca caagcgtggg ctgaaacggg cgggatgggc  120
ctcgctgtaa gacaagcgcc acttatcatc ccgcttaaag ctacttctac tcccgtctct  180
attaagcaat acccaatgtc ccaggaagct cgtttgggca ttaagcctca tatacaaagg  240
ctactcgatc agggcatact tgttccctgc caatcaccgt ggaatacgcc cctattacca  300
gttaagaagc ctgggactaa cgactatcgc cccgtacagg atctacgtga ggtgaacaag  360
cgtgtagagg acatccatcc gaccgttccg aatccataca atttgctttc tggattacct  420
ccaagtcatc aatggtacac cgtgctggat ctcaaggacg cattcttctg tctaagatta  480
catcctacta gccagccact tttcgcattc gagtggcgag atcccgagat gggaatttcg  540
ggccagctta catggacgag gcttcctcaa ggcttcaaga actctcctac cttgttcaat  600
gaggctctac accgcgacct cgcagacttc cggatacaac atccggacct catactccta  660
caatatgtgg acgatctatt gctggccgcg acgagcgaat tggattgtca gcaaggaacc  720
cgcgccttgt tacaaacgtt ggggaacttg ggtatcgag catcagccaa gaaggcacaa  780
atctgccaga aacaagtgaa gtatttgggg tacttactga aagaggggca acgatggttg  840
accgaagctc gcaaggaaac tgttatgggc cagccgacac ctaagactcc aagacagctc  900
cgagagttcc tcggcaaggc tgggttctgt cgcctattca ttcctgggtt tgccgaaatg  960
gctgctcctc tgtacccgtt gaccaaaccg ggaaccttgt tcaattgggg accagatcaa  1020
cagaaggcgt atcaggagat caagcaagcg ctgttgactg cgcctgcgtt gggcttgccg  1080
gatttgacaa aaccctttga acttttcgtt gacgagaaac aaggttacgc gaagggagtt  1140
cttacacaaa agctgggccc ctggagacga cctgttgctt atcttagcaa gaagttagat  1200
cccgttgctg cggggtggcc gccctgcttg aggatggttg ccgccattgc ggttctgact  1260
aaagatgcgg gcaaattgac gatgggccag ccgctcgtaa ttctcgcccc gcacgcagtc  1320
gaagccctag tgaagcagcc tcctgaccgt tggctctcca acgctcggat gacccactat  1380
caagcgctgc tgttggatac cgatagagtt caattcgggc cggtcgtagc gctcaatccc  1440
gcaacgctat taccccctgcc tgaggaggga ctacaacata actgcttgga tattctagcg  1500
gaggctcatg ggacaagacc tgacttgaca gatcagccct tgccagatgc cgaccacaca  1560
tggtacactg acggctcatc acttctacaa gagggccaaa ggaaggccgg agctgcggtg  1620
acgactgaaa ccgaggtgat ctgggcaaag gctttacccg ctggaacttc tgctcagcgc  1680
gcagagctta tcgccctaac tcaagccctg aaaatggctg agggcaagaa gttgaatgtc  1740
tataccgatt cacggtatgc tttcgctacc gcccacattc atggagaaat ctatcggcga  1800
cgaggttggc tcacgtctga ggggaaggag attaagaaca aggacgaaat cttagctctg  1860
ctgaaagctc tattcttacc caaacgcttg tcgattatcc actgcccagg acaccaaaag  1920
ggccattccg ccgaagccag gggaaaccgt atggccgatc aagctgcccg gaaggctgca  1980
ataaccgaaa cccctgacac ttccacgcta ctcatcgaaa actctagccc a           2031

SEQ ID NO: 165         moltype = DNA  length = 2031
FEATURE                Location/Qualifiers
misc_feature           1..2031
                       note = MMLV_DO2
source                 1..2031
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 165
accttgaata tcgaagatga gtacagactc catgagacca gcaaggagcc tgatgtgagc  60
ctgggaagta catggctatc tgactttcct caggcttggg cggagaccgg tggcatgggc  120
ctggctgtcc ggcaagcacc cctaatcatc ccgttgaaag caacgtccac acccgtatct  180
atcaagcagt acccgatgag ccaagaggca cggctaggga ttaagccgca catccaaaga  240
ttgttagacc aggggatact cgttccctgt caatctccgt ggaacacacc gttactgcct  300
gttaagaagc ccggcacaaa tgattatcgg cctgtacaag acctgcgcga agtgaacaaa  360
cgagtggaag atattcaccc caccgtgccg aatccttaca acttgctaag cggtttacca  420
cccagtcacc agtggtacac cgtcttggac ctcaaagacg cattcttctg cttgcggttg  480
catcctacga gtcagccgct ctttgccttt gaatggagag atcccgagat gggcataagt  540
ggtcagctta catggacgag gctacctcag ggattcaaga actcgcctac cttattcaat  600
```

-continued

```
gaggctctac acagggatct tgctgacttt cggatacaac accctgactt aattctgcta  660
caatatgttg acgatctgct gcttgcggcg accagtgaac tcgactgcca acaagggaca  720
agagccttgt tacagacact tggcaatctc ggctaccgcg cgtctgctaa gaaggcacag  780
atttgccaaa agcaagtgaa gtatctcggt tacctgctga aagagggaca aagatggttg  840
accgaagcca ggaaggaaac tgtcatgggg caacccaccc ctaaaacccc aagacagctt  900
agagagttcc tcgggaaagc tgggttctgc cgcctgttca ttcccggttt cgctgaaatg  960
gctgcacctc tatacccact gaccaaaccg ggcacgctat tcaattgggg accggatcaa  1020
caaaaggctt accaggagat caagcaagct ctattgacag ctccggctct gggtttaccc  1080
gacttgacta aacccttcga gttgttcgtt gacgagaaac agggctacgc gaaaggcgta  1140
ctgacgcaga aactgggccc gtggcggcga cccgttgctt acctttccaa gaagctggac  1200
cctgttgccg ctggctggcc gccctgcctt cgcatggttg cggccattgc tgtccttacg  1260
aaagatgctg ggaaactaac tatgggtcag ccgctcgtaa ttctcgcgcc acacgctgtc  1320
gaggctctag tcaaacaacc tcctgaccgc tggctgtcga atgcacggat gacacattac  1380
caggcgttgc tattggacac tgaccgagta caatttggac cagtagtggc tttaaatccc  1440
gcgacactcc ttccccttcc tgaggagggt ttgcaacaca actgtcttga tatacttgct  1500
gaggcccacg gaacaagacc cgatctcaca gatcaacccc tccccgacgc agatcacacc  1560
tggtacaccg acggttcaag cctgttacag gaaggccagc gaaaggccgg agcagccgtg  1620
acaaccgaaa ccgaagttat ctgggcaaag gccctaccag ccgggacaag cgcgcagagg  1680
gccgagctga tcgcgctcac acaagcactc aaaatggccg aaggcaagaa gctcaatgtt  1740
tacactgatt cccgctacgc attcgctacc gctcatattc acggagaaat ctacaggagg  1800
cgagggtggc ttactagcga aggaaaggag attaagaaca aagacgagat cctggcacta  1860
ttgaaagcct tgttcctgcc aaagcgctta tctatcattc actgtccggg ccaccagaag  1920
ggccacagtg ctgaggctcg ggcaatcgg atggcagatc aggccgcacg aaaagccgca  1980
attactgaga cacctgacac atctaccttg ctgattgaga actcgtctcc g           2031

SEQ ID NO: 166         moltype = DNA  length = 2073
FEATURE                Location/Qualifiers
source                 1..2073
                       mol_type = other DNA
                       organism = Soybean chlorotic mottle virus
SEQUENCE: 166
aatactgaaa ttgtccaaaa acaccgagtt ttaaccaaag gtaaccctaa tgttactttc  60
ataaaagtta gtataggcaa aagaaatttc ttggcttata ttgatactgg agcaactctg  120
tgctttggaa aaagaaaaat ttcaaataat tgggaaattt taaaacaacc aaaagaaatt  180
atcattgcag ataaatcaaa acactatatt agagaagcta tttctaatgt gtttttaaaa  240
atcgaaaata aagaattctt aatccctatc atatatttac atgattcagg attagattta  300
attataggaa acaatttcct aaaattatac caacctttta ttcagagatt ggaaacaatt  360
gaattaagat ggaaaaatct taataaccca aaagaatctc aaatgatttc aaccaagatt  420
cttacaaaaa atgaagtatt aaaactttca tttgaaaaaa ttcatatttg tttagaaaaa  480
tatttatttt tcaaaacaat tgaagaacaa ctcgaagaag tatgttcaga acatccactg  540
gatgaaacaa aaaataaaaa tggtctttta atagaaataa gacttaaaga cccattacaa  600
gaaataaatg tcacaaatag aattccatat acaataagag atgtacaaga attcaaggag  660
gaatgtgaag acctcttaaa aaagggctta attcgagaat ctcaaagtcc acacagtgca  720
ccggcattct atgtcgaaaa tcacaatgaa atcaagcgtg gaaaaagacg catggtaatt  780
aattacaaaa aaatgaatga agccacaatt ggcgattcat ataagttacc aagaaaagat  840
tttattctgg aaaaaataaa aggatcttta tggttttcaa gcttggatgc taaatctgga  900
tactaccagc taaggctcca tgaaaataca aagcctctaa cagctttttc atgtccacct  960
cagaaacatt acgaatggaa tgttttaagt tttggactta aacaagcacc atctatatat  1020
caaagattta tggatcaatc cctcaaggga cttgaacata tatgtttggc atatattgat  1080
gacatcctga tctttacaaa aggatctaaa gaacaacatg taaatgatgt tcggattgtt  1140
ttgcaaagaa tcaaagaaaa aggaattatt atttctaaga aaaaatcaaa actgattcaa  1200
caggaaatcg aatatctcgg tttaaaaata caagggaatg gagaaattga tttatcacct  1260
catacccaag aaaaaattct tcaatttcct gatgaattag aagatagaaa acaaatacag  1320
cgttttcttg gctgtattaa ttacattgca aatgaaggat tttcaaaaa tcttgctcta  1380
gaaagaaagc accttcaaaa gaaaatttct gttaaaaacc cctggaaatg ggatacaata  1440
gatacaaaaa tggttcagtc cataaaaggc aaaattcaaa gcctaccaaa attatataat  1500
gcatcgattc aagacttttt aatagtcgag acagatgcat cgcaacactc ctggagtgga  1560
tgtttgcgag ctttacccaa gggaaagcaa aaaatcggac tcgatgaatt cgggataccg  1620
acagctgacc tctgcacagg tagcagttca gcttcaagcg ataattcgcc agctgagatt  1680
gacaaatgtc attcagccag taaacaggac actcatgtgg ccagtaaaat aaagaaactc  1740
gaaaacgagc ttctactttg caaatatgtt tcaggtacct tcacagatac ggaaacaaga  1800
taccctatag cagaactgga ggttcttgct ggagtaaaag tcctagaaaa atggagaatc  1860
gacctcctac aaacgaggtt cctcctccgc actgacagca agtactttgc aggtttttgt  1920
aggtacaaca tcaagacaga ctaccggaac ggacgtctaa tcaggtggca actacggtta  1980
caagcctatc aaccgtacgt ggaattaatc aaatcagaaa ataacccatt cgcagatacg  2040
cttacgcgag aatggagcaa gccatcaagc agt                               2073

SEQ ID NO: 167         moltype = DNA  length = 2073
FEATURE                Location/Qualifiers
misc_feature           1..2073
                       note = SbCMV_DO1
source                 1..2073
                       mol_type = other DNA
                       organism = synthetic construct
SEQUENCE: 167
aacactgaga ttgttcaaaa acatagagtg cttaccaaag gaaatccaaa tgttacattt  60
```

-continued

```
ataaaagtgt ccatagggaa gagaaatttt ttagcttata ttgacactgg agccacactc  120
tgttttggaa aaaggaaaat atcaaataac tgggaaatcc ttaagcaacc caaagaaatc  180
attatcgctg ataagtcaaa acactacatc agagaagcta taagtaacgt attcctgaaa  240
attgaaaaca aggagttctt gatacccatt atatatcttc atgattcagg gttggatttg  300
attattggga ataacttcct gaagctttat caaccattta ttcaaagact tgaaactatc  360
gaactcaggt ggaaaaactt gaacaatccc aaagagtctc aaatgattag cactaaaatt  420
cttacgaaaa atgaagttct taagctgagt tttgagaaga ttcatatttg tctcgaaaaa  480
taccttttct ttaaaaccat cgaggaacaa cttgaggagg tttgttctga acatccactt  540
gatgagacaa agaacaagaa tggtcttttg attgagatac gtctgaaaga tcctctgcag  600
gagattaacg tcacaaatag gattccatat accattagag atgtacagga attcaaggaa  660
gaatgtgaag atttacttaa gaagggtctc attcgtgaat cacaatctcc ccacagtgca  720
cccgcatttt acgttgaaaa tcataatgaa attaagagag gcaagcgtag aatggttatc  780
aactacaaga agatgaatga agcaaccata ggagatagct acaaactccc gcgaaaggat  840
tttatcttag agaagataaa gggcagtttg tggttttcaa gtttagatgc aaaatcaggt  900
tattatcagc ttcgcttaca tgagaacaca aagcctctca ctgctttctc ttgccctcct  960
caaaaacatt atgaatggaa tgtgttgagt ttcggtctaa aacaggcacc ttcgatttac  1020
cagcgcttca tggaccagtc cttaaaggga ttagagcaca tttgcttggc atatatagat  1080
gatatcttaa tctttactaa aggctcaaag gaacagcatg tcaatgatgt tcggattgtc  1140
ctgcaaagaa taaaagagaa aggaatcata atatctaaaa aaaaatcaaa attgattcag  1200
caagagattg aatatctagg attgaaaatt caaggtaatg gtgaaattga cctctcacca  1260
catactcaag aaaagatcct acagttccct gatgaactgg aggatagaaa acaaatacag  1320
aggtttctag gttgcattaa ttacattgcg aacgaaggat ttttcaaaaa tcttgcccta  1380
gagagaaagc acttgcagaa gaagatttcc gtgaagaatc catggaagtg ggatacaata  1440
gacacaaaaa tggtgcaatc aatcaagggc aaaattcaat ccctgccgaa gctctacaat  1500
gcaagtattc aggatttcct aattgtagag actgacgcct cgcaacattc ttggtctggg  1560
tgtttgcggg ctcttccaaa gggcaagcag aaaatcggtc tggacgaatt tgggattcca  1620
acggcagatt tatgtactgg tagctccagt gcttcctctg ataattctcc tgctgagatc  1680
gacaagtgcc actcagcctc gaagcaggat acacacgtcg cctctaaaat aaagaaactt  1740
gagaatgagt tacttttgtg caagtatgtt tcagggactt tcacggacac cgagactagg  1800
tatcctatag ctgaactcga ggtgttggcg ggtgttaaag ttttggaaaa atggaggata  1860
gacttgttgc aaacacgatt tctacttagg acagattcca aatattttgc tggattttgt  1920
agatacaaca ttaagactga ttatcggaac gggaggctca taagatggca attgcgcctt  1980
caagcttacc agccttatgt ggaactgatc aagagtgaaa ataatccttt tgcagacacc  2040
ctaacacgag agtggagcaa accatcttct agc                                2073

SEQ ID NO: 168         moltype = DNA  length = 2034
FEATURE                Location/Qualifiers
source                 1..2034
                       mol_type = other DNA
                       organism = Cauliflower mosaic virus
SEQUENCE: 168
gatcatctac ttctgaagac tcagactcag actgagcagg tgatgaacgt caccaatccc  60
aattcgatct acatcaaggg aagactctac ttcaagggat acaagaagat agaacttcac  120
tgtttcgtag acacgggagc aagcctatgc atagcatcca agttcgtcat accagaagaa  180
cattgggtca atgcagaaag accaattatg gtcaaaatag cagatggaag ctcaatcacc  240
atcagcaaag tctgcaaaga catagacttg atcatagccg gcgagatatt cagaattccc  300
accgtctatc agcaagaaag tggcatcgat ttcattatcg gcaacaactt ctgtcagctg  360
tatgaaccat tcatacagtt tacggataga gttatcttca caaagaacaa gtcttatcct  420
gttcatattg cgaagctaac cagagcagtg cgagtaggca ccgaaggatt tcttgaatca  480
atgaagaaac gttcaaaaac tcaacaacca gagccagtga acatttctac aaacaagata  540
gaaaatccac tagaagaaat tgctattctt tcagagggga ggaggttatc agaagaaaaa  600
ctctttatca ctcaacaaag aatgcaaaaa atcgaagaac tacttgagaa agtatgttca  660
gaaaatccat tagatcctaa caagactaag caatggatga aagcttctat caagctcagc  720
gacccaagca aagctatcaa ggttaaaccc atgaagtata gcccaatgga tcgcgaagaa  780
tttgacaagc aaatcaaaga attactggac ctaaaagtca tcaagcccag taaaagccct  840
cacatggcac cagccttctt ggtcaacaat gaagccgaga agcgaagagg aaagaaacgt  900
atggtagtca actacaaagc tatgaacaaa gctactgtag gagatgccta caatcttccc  960
aacaaagacg agttacttac actcattcga ggaaagaaga tcttctcttc cttcgactgt  1020
aagtcaggat tctggcaagt tctgctagat caagaatcaa gacctctaac ggcattcaca  1080
tgtccacaag gtcactacga atggaatgtg gtccctttcg gcttaaagca agctccatcc  1140
atattccaaa gacacatgga cgaagcattt cgtgtgttca gaaagttctg ttgcgtttat  1200
gtcgacgaca ttctcgtatt cagtaacaac gaagaagatc atctacttca cgtagcaatg  1260
atcttacaaa agtgtaatca acatggaatt atcctttcca agaagaaagc acaactcttc  1320
aagaagaaga taaacttcct tggtctagaa atagatgaag gaacacataa gcctcaagga  1380
catatcttgg aacacatcaa caagttcccc gataccctţg aagacaagaa gcaacttcag  1440
agattcttag gcatactaac atatgcctcg gattacatcc cgaagctagc tcaaatcaga  1500
aagcctctgc aagccaagct taaagaaaac gttccatgga gatggacaaa agaggatacc  1560
ctctacatgc aaaaggtgaa gaaaaatctg caaggatttc ctccactaca tcatccctta  1620
ccagaggaga agctgatcat cgagaccgat gcatcagacg actactgggg aggtatgtta  1680
aaagctatca aaattaacga aggtactaat actgagttaa tttgcagata cgcatctgga  1740
agctttaaag ctgcagaaaa gaattaccac agcaatgaca aagagacatt ggcggtaata  1800
aatactataa agaaatttag tatttatcta actcctgttc attttctgat taggacagat  1860
aatactcatt tcaagagttt cgttaatctc aattacaaag gagattcgaa acttggaaga  1920
aacatcagat ggcaagcatg gcttagccac tattcatttg atgttgaaca cattaaagga  1980
accgacaacc actttgcgga cttcctttca agagaattca ataaggttaa ttcc        2034
```

-continued

```
SEQ ID NO: 169          moltype = DNA  length = 2034
FEATURE                 Location/Qualifiers
misc_feature            1..2034
                        note = CaMV_DO1
source                  1..2034
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 169
gatcaccttc tgcttaagac acagacacaa actgaacaag ttatgaatgt gaccaacccg  60
aattccattt acattaaggg acgactctac tttaaggggt ataagaaaat agaattacat  120
tgtttcgtcg acactggagc atctctctgc atagcgtcca agtttgttat tcctgaggaa  180
cattgggtaa atgcagaaag gccgattatg gttaaaattg cagatggtag cagcatcaca  240
atctcaaaag tttgcaagga cattgacttg atcattgctg gtgagatttt cagaattcct  300
acagtgtatc aacaagaatc cggcattgat tttataattg gtaataactt ttgtcaactt  360
tatgagccct tcatacaatt tacagatcga gtcattttta ctaaaaacaa gagttaccct  420
gttcacattg caaaactcac tcgtgccgtg agagttggaa cggaaggatt tctagaatct  480
atgaagaaga gatcgaaaac tcagcaacca gaacccgtta atatttctac aaataagatt  540
gaaaatccat tagaggaaat agccatcttg tccgaaggcc ggcggttgag tgaagaaaag  600
ttgtttatca cgcagcagag aatgcaaaaa atagaggagc ttctcgaaaa ggtttgttct  660
gagaatcctt tggatccaaa taaaacaaaa caatggatga aagctagtat aaagctttca  720
gacccatcaa aggcaattaa ggtgaagcca atgaaatata gccccatgga tagggaggag  780
tttgacaagc aaattaagga gctactcgat ctgaaagtaa taaaaccttc taaatcgcct  840
cacatggctc cagcattcct ggttaacaac gaggctgaaa agcgcagagg aaaaaaaaga  900
atggtggtga actacaaagc aatgaataag gctactgttg gagatgctta taatcttcct  960
aataaagatg agctcttgac cttaattaga gggaagaaaa ttttctcctc atttgattgt  1020
aaatcaggat tttggcaagt gttgctggat caagagtctc gtccactgac cgcctttacg  1080
tgccctcaag gacattatga atggaatgtc gtaccatttg gtctcaagca agcaccttct  1140
attttccaga ggcatatgga tgaagcattt agagtgttta ggaaattctg ctgtgtttat  1200
gtggatgata tattggtatt ctcaaataat gaggaagacc atttgctgca tgttgccatg  1260
attcttcaga agtgcaatca acatggaatc atcttatcca agaagaaggc tcagttgttc  1320
aagaagaaga taaatttttt gggtctcgag attgatgagg ggacacataa gcctcagggt  1380
catatactag aacatatcaa caagtttcca gacactttgg aagacaaaaa gcagttgcaa  1440
aggttccttg ggattctgac ttatgcttca gattatatac caaagcttgc tcaaataaga  1500
aaacccettc aggcgaagct caaagaaaac gttccttgga ggtggactaa ggaggatacc  1560
ttatacatgc agaaagtcaa gaaaaacctc cagggtttcc caccgctcca ccatccttta  1620
cctgaagaaa aactaattat cgagacagat gcttctgatg actactgggg cggcatgttg  1680
aaggccatca aaatcaatga agggaccaat actgagctca tttgtcgata tgcaagcgga  1740
agttttaaag cagctgagaa aaattatcat agtaatgata aagagactct agccgttatt  1800
aacaccataa agaaattctc tatatatctt acccccgtcc actttttaat caggacagac  1860
aacactcact tcaaatcatt tgtgaacctg aattacaagg gtgatagtaa acttggccgt  1920
aacatacgct ggcaggcttg gttgagccac tactcttttg atgtagaaca cattaaagga  1980
acagataatc attttgctga tttcctttct cgcgagttca acaaagtaaa ttca        2034

SEQ ID NO: 170          moltype = AA  length = 683
FEATURE                 Location/Qualifiers
source                  1..683
                        mol_type = protein
                        organism = Moloney Murine Leukemia Virus
SEQUENCE: 170
TLNIEDEHRL HETSKEPDVS LGSTWLSDFP QAWAETGGMG LAVRQAPLII PLKATSTPVS  60
IKQYPMSQEA RLGIKPHIQR LLDQGILVPC QSPWNTPLLP VKKPGTNDYR PVQDLREVNK  120
RVEDIHPTVP NPYNLLSGLP PSHQWYTVLD LKDAFFCLRL HPTSQPLFAF EWRDPEMGIS  180
GQLTWTRLPQ GFKNSPTLFD EALHRDLADF RIQHPDLILL QYVDDLLLAA TSELDCQQGT  240
RALLQTLGNL GYRASAKKAQ ICQKQVKYLG YLLKEGQRWL TEARKETVMG QPTPKTPRQL  300
REFLGTAGFC RLWIPGFAEM AAPLYPLTKT GTLFNWGPDQ QKAYQEIKQA LLTAPALGLP  360
DLTKPFELFV DEKQGYAKGV LTQKLGPWRR PVAYLSKKLD PVAAGWPPCL RMVAAIAVLT  420
KDAGKLTMGQ PLVILAPHAV EALVKQPPDR WLSNARMTHY QALLLDTDRV QFGPVVALNP  480
ATLLPLPEEG LQHNCLDILA EAHGTRPDLT DQPLPDADHT WYTDGSSLLQ EGQRKAGAAV  540
TTETEVIWAK ALPAGTSAQR AELIALTQAL KMAEGKKLNV YTDSRYAFAT AHIHGEIYRR  600
RGLLTSEGKE IKNKDEILAL LKALFLPKRL SIIHCPGHQK GHSAEARGNR MADQAARKAA  660
ITETPDTSTL LIENSSPNSR LIN                                          683

SEQ ID NO: 171          moltype = AA  length = 717
FEATURE                 Location/Qualifiers
source                  1..717
                        mol_type = protein
                        organism = Moloney Murine Leukemia Virus
SEQUENCE: 171
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
```

-continued

```
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPSGGS KRTADGSEFE PKKKRKV     717

SEQ ID NO: 172         moltype = AA  length = 19
FEATURE                Location/Qualifiers
source                 1..19
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 172
MKRTADGSEF ESPKKKRKV                                               19

SEQ ID NO: 173         moltype = AA  length = 21
FEATURE                Location/Qualifiers
source                 1..21
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 173
SGGSKRTADG SEFEPKKKRK V                                            21

SEQ ID NO: 174         moltype = AA  length = 7
FEATURE                Location/Qualifiers
source                 1..7
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 174
PKKKRKV                                                            7

SEQ ID NO: 175         moltype = AA  length = 175
FEATURE                Location/Qualifiers
source                 1..175
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 175
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFR       175

SEQ ID NO: 176         moltype = AA  length = 1198
FEATURE                Location/Qualifiers
source                 1..1198
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 176
CINENLTRYI SNMDIFEKVD AIFDKHEVQE IKEKILNSDY DVEDFFEGEF FNFVLTQEGI  60
DVYNAIIGGF VTESGEKIKG LNEYINLYNQ KTKQKLPKFK PLYKQVLSDR ESLSFYGENQ  120
TTQKGQKNSR ERMKRIEEGI KELGSQILKE HPVENTQLQN EKLYLYYLQN GRDMYVDQEL  180
DINRLSDYDV DHIVPQSFLK DDSIDNKVLT RSDKNRGKSD NVPSEEVVKK MKNYWRQLLN  240
AKLITQRKFD NLTKAERGGL SEGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE  300
YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK  360
KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND  420
AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY  480
VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID  540
KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN  600
DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV  660
DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS  720
LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF  780
KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT  840
DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG  900
FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN  960
GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK  1020
NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD  1080
IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ  1140
ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKH    1198

SEQ ID NO: 177         moltype = AA  length = 455
FEATURE                Location/Qualifiers
source                 1..455
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 177
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
```

-continued

```
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNK                             455

SEQ ID NO: 178         moltype = AA  length = 918
FEATURE                Location/Qualifiers
source                 1..918
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 178
NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH  60
LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG  120
SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF  180
VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG  240
SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE  300
DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY  360
REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE  420
NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS  480
EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI  540
VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK  600
ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA  660
LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS  720
FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC  780
LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL  840
ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK  900
IAISNKEWLE YAQTSVKH                                                918

SEQ ID NO: 179         moltype = AA  length = 551
FEATURE                Location/Qualifiers
source                 1..551
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 179
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD A                                                       551

SEQ ID NO: 180         moltype = AA  length = 822
FEATURE                Location/Qualifiers
source                 1..822
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 180
DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF  60
ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ  120
NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL  180
PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY  240
PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN  300
KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA  360
NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP  420
YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA  480
RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF  540
EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP  600
STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY  660
SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS  720
FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN  780
IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KH                     822

SEQ ID NO: 181         moltype = AA  length = 586
FEATURE                Location/Qualifiers
source                 1..586
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 181
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
```

-continued

```
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFV                 586

SEQ ID NO: 182          moltype = AA  length = 787
FEATURE                 Location/Qualifiers
source                  1..787
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 182
LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD  60
HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK  120
YAKCLQKIDK DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF  180
KKGDMFNLND CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS  240
FESASKKEVD KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG  300
AELFMRRASL KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA  360
INKCPKNIFK INTEVRVLLK HDDNPYVIGI ARGERNLLYI VVVDGKGNIV EQYSLNEIIN  420
NFNGIRIKTD YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV  480
IALEDLNSGF KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE  540
SFKSMSTQNG FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED  600
LFEFALDYKN FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK  660
YGINYQQGDI RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF  720
YDSRNYEAQE NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY  780
AQTSVKH                                                            787

SEQ ID NO: 183          moltype = AA  length = 695
FEATURE                 Location/Qualifiers
source                  1..695
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 183
MGSKLEKFTN CYSLSKTLRF KAIPVGKTQE NIDNKRLLVE DEKRAEDYKG VKKLLDRYYL  60
SFINDVLHSI KLKNLNNYIS LFRKKTRTEK ENKELENLEI NLRKEIAKAF KGNEGYKSLF  120
KKDIIETILP EFLDDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN  180
LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA  240
IIGGFVTESG EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGENQTTQKG  300
QKNSRERMKR IEEGIKELGS QILKEHPVEN TQLQNEKLYL YYLQNGRDMY VDQELDINRL  360
SDYDVDHIVP QSFLKDDSID NKVLTRSDKN RGKSDNVPSE EVVKKMKNYW RQLLNAKLIT  420
QRKFDNLTKA ERGGLSEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG  480
IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF  540
SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS LKKNDAVVAI  600
MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP  660
YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYG                             695

SEQ ID NO: 184          moltype = AA  length = 678
FEATURE                 Location/Qualifiers
source                  1..678
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 184
SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE  60
DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY  120
REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE  180
NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS  240
EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI  300
VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK  360
ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA  420
LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS  480
FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC  540
LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL  600
ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK  660
IAISNKEWLE YAQTSVKH                                                678

SEQ ID NO: 185          moltype = DNA  length = 2538
FEATURE                 Location/Qualifiers
source                  1..2538
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 185
atgaaacgga cagccgacgg aagcgagttc gagtcaccaa agaagaagcg gaaagtcaca  60
cttaatattg aggatgaaca tagattgcac gagacctcta aggaacctga tgtttctctt  120
ggatcaactt ggttgtcaga tttcccacaa gcatgggcag agaccggagg tatgggtctt  180
gctgttaggc aggcaccact tattattcct ttgaaggcaa cctctactcc tgtgtcaatt  240
```

-continued

```
aagcaatatc caatgtctca ggaagctagg cttggaatta agcctcacat tcaaagactt  300
ttggatcagg gtattttggt gccatgtcaa tcaccttgga acacaccact tttgcctgtt  360
aagaagcctg gaactaatga ttacagacca gtgcaagatt tgagggaggt taacaagaga  420
gtggaagata ttcacccaac tgttccaaac ccttataatc ttttgtctgg attgccacct  480
tcacatcaat ggtacactgt gcttgatttg aaggatgcat tttctgcct taggttgcat  540
ccaacatctc agcctctttt tgctttcgag tggagagatc ctgaaatggg aatttctggt  600
caacttacat ggaccaggtt gcctcagggt ttcaagaact caccaacctt gtttaatgag  660
gcacttcaca gggatttggc tgattttagg attcaacatc ctgatcttat ccttttgcag  720
tatgttgatg atcttttgct tgctgcaact tctgaattgg attgtcaaca gggaactagg  780
gcattgcttc aaacacttgg aaatttgggt tacagagctt cagcaaagaa ggctcagatt  840
tgccaaaagc aggttaagta tcttggatac ttgcttaagg aaggacaaag gtggttgacc  900
gaggctagaa aggaaactgt gatgggtcaa ccaacaccta agacccctag gcagcttaga  960
gagttcttgg gaaaggcagg tttttgtagg cttttcattc caggatttgc tgaaatggct  1020
gcaccacttt atcctttgac caagcctgga actttgttta actggggtcc agatcaacag  1080
aaggcatacc aagaaattaa gcaggctttg cttactgctc cagcacttgg tttgcctgat  1140
cttacaaagc catttgagtt gttcgttgat gaaaagcaag gatatgcaaa gggtgtgctt  1200
acccagaagt tgggaccttg gagaaggcct gttgcttacc tttctaagaa acttgatcca  1260
gtggctgcag gttggccacc ttgtcttaga atggttgctg caattgcagt gcttacaaag  1320
gatgctggaa agttgactat gggacaacct cttgttattt tggcaccaca cgctgttgag  1380
gcacttgtga agcagccacc tgataggtgg ttgtcaaacg caagaatgac ccattatcaa  1440
gctcttcttt tggatactga tagggtgcag ttcggtcctg ttgtggcttt gaatccagca  1500
acacttttgc cacttcctga ggaaggattg caacacaact gccttgatat tttggctgag  1560
gcacatggta caagacctga tcttaccgat cagccattgc ctgatgctga tcacacttgg  1620
tacacagatg gatcttcact tttgcaagaa ggacagagga aggctggtgc tgcagttact  1680
acagagactg aagtgatttg ggctaaggca cttccagctg gaacatctgc tcaaagagca  1740
gagcttattg ctttgaccca ggcacttaag atggctgaag gaaagaagtt gaacgtttac  1800
actgattcta ggtatgcttt cgcaacagct catattcacg gagaaatcta tagaaggaga  1860
ggatggttga catcagaggg aaaggaaatt aagaacaagg atgaaattct tgcacttttg  1920
aaggctcttt ttcttcctaa gagattgtct attattcatt gcccaggaca ccaaaagggt  1980
cattcagcag aagctagggg aaatagaatg gctgatcagg ctgcaagaaa ggctgcaatt  2040
actgagacac ctgatacctc tactcttttg atcgaaaact cttcaccaaa ttcaaggctt  2100
attaactctg gtggttcaag cggaggctca tctggcagcg aaactccggg cacttccgag  2160
tcagcttgcc tgtcctacga gacagagatc ctgacagtgg agtatggcct gctgccaatc  2220
ggcaagatcg tggagaagag gatcgagtgt accgtgtact ctgtggataa caatggcaac  2280
atctatacac agcccgtggc acagtggcac gataggggag agcaggaggt gttcgagtat  2340
tgcctggagg acggcagcct gatcagggca accaaggacc acaagttcat gacagtggat  2400
ggccagatgc tgcccatcga cgagattttc gagcgggagc tggacctgat gagagtggat  2460
aacctgccta atagcggagg cagtaaaaga acagcagacg ggagtgagtt tgagcccaag  2520
aaaaagagaa aggtgtaa                                                2538

SEQ ID NO: 186          moltype = DNA  length = 4155
FEATURE                 Location/Qualifiers
source                  1..4155
                        mol_type = other DNA
                        organism = synthetic construct
SEQUENCE: 186
atgaaacgga cagccgacgg aagcgagttc gagtcaccaa agaagaagcg gaaagtcatc  60
aagattgcta cacggaaata cctgggaaag cagaacgtgt acgacatcgg cgtgggcgag  120
ccacacaact tcgccctgaa gaatggcttt atcgccagca attgcgcaac tcctgagtct  180
agcggcgggt cgtcaggagg gtcttccaag ctggagaagt ttacaaactg ttacagcctc  240
tccaaaaccc tcaggtttaa agcgatcccg gtgggcaaga cccaggagaa catcgacaac  300
aagaggctcc tggtggaaga cgagaagcgc gccgaagact acaagggcgt gaagaagctg  360
ctcgataggt actacctcag ctttattaac gacgtgctgc acagcatcaa actcaagaat  420
ctcaacaact acatctccct cttccgcaaa aagacccgca ccgagaagga gaacaaggag  480
ctggagaacc tggagatcaa cctccgcaag gaaatcgcca aagcgttcaa gggcaatgaa  540
gggtacaaga gcctcttcaa gaaagacatc atcgaaacta tcctcccaga gtttctcgat  600
gacaaggacg agatcgcgct ggtgaactcc tttaacgggt tcacaaccgc gtttaccggc  660
ttctttgata acagggaaaa tatgttctcc gaggaggcca agtccaccag catcgccttc  720
aggtgtatca acgagaacct cacccgctac atttccaata tggacatttt cgagaaggtg  780
gatgcgatct tcgataagca cgaggtgcag gagatcaaag agaagattct caattccgat  840
tatgacgtcg aggatttctt cgaaggggag ttctttaatt ttgtgctcac acaagagggc  900
attgacgtgt acaacgcgat tatcgggggc ttcgtcacag agtccgggga gaagattaag  960
gggctgaatg agtacatcaa tctgtacaat cagaagacca agcagaaact gccgaaattc  1020
aagccgctct acaagcaagt cctgtccgat agggaaagcc tctccttcta cggcgagggc  1080
tataccagcg acgaggaggt gctggaagtc ttccgcaaca cactgaataa gaatagcgag  1140
attttctcct ccatcaagaa gctcgagaag ctctttaaga actttgacga gtacagctcc  1200
gccgggattt tcgtgaagaa cgggccggcg atcagcacca tctccaagga catctttggc  1260
gagtggaacg tcatcaggga caagtggaac gccgagtacg acgacatcca cctgaagaag  1320
aaggcggtgg tgaccgagaa gtatgaggac gatcgcagga agtccttcaa aaaaatcggc  1380
tccttcagcc tcgaacagct ccaggagtat gccgatgcgg atctgtccgt cgtcgagaag  1440
ctgaaggaaa tcatcattca gaaggtcgac gagatctata aagtgtacgg gtccagcgag  1500
aagctgttcg acgccgactt tgtgctcgag aagtccctca aaaagaatga cgccgtggtg  1560
gccattatga aagacctgct cgactccgtg aagtccttcg aaaattacat taaagcgttc  1620
tttggggagg ggaaggaaac taacagggat gagtccttct atggcgactt tgtcctcgcg  1680
tacgacatcc tgctgaaggt cgaccacatt tacgacgcga tccgcaacta cgtgacacag  1740
aagccgtact ccaaagacaa gttcaagctg tacttccaga acccgcaatt tatgggggc  1800
```

-continued

```
tgggacaagg ataaagagac agactaccgc gcgacaattc tccgctatgg ctccaaatac  1860
tatctggcca tcatggacaa gaagtacgcg aagtgcctgc agaagatcga caaagacgac  1920
gtcaatggca actatgaaaa gatcaactac aagctgctgc cgggcccgaa caagatgctc  1980
ccgaaggtgt tcttcagcaa gaagtggatg gcctactaca atccaagcga ggatattcag  2040
aaaatctata aaaacgggac cttcaagaag ggggacatgt ttaacctcaa cgactgccac  2100
aagctcattg atttcttcaa ggatagcatt tcccgctacc cgaaatggtc caatgcgtac  2160
gattttaact tctccgagac agaaaagtac aaagacatcg cgggctttta cagggaggtg  2220
gaggagcaag ggtataaagt ttcttttgaa tccgcgagca agaaggaagt cgacaagctc  2280
gtcgaggagg gcaagctcta catgttccaa atttataaca aggacttttc cgacaagagc  2340
catgggaccc caaacctcca caccatgtac ttcaaactgc tctttgacga gaacaaccac  2400
gggcaaatca ggctgagcgg cggcgccgaa ttattcatgc gcagggcctc cctcaagaag  2460
gaagagctgg tcgtccatcc agccaattcc ccgatcgcga acaagaaccc ggacaatccg  2520
aaaaagacca ccaccctgtc ctacgacgtc tacaaggaca aacgcttcag cgaagaccag  2580
tacgaattac acatcccaat tgcgattaat aagtgcccaa agaatatctt caaaattaat  2640
acagaggtca gggtgctgct caaacacgac gacaatccgt atgtcatcgg cattgacagg  2700
ggcgagcgca atctgctcta tatcgtggtc gtggatggga agggcaatat tgtggagcag  2760
tactccctga acgagattat caacaacttc aatgggatta ggattaagac cgactatcac  2820
agcctgctcg acaagaaaga aaaagagagg tttgaggccc gccaaaactg gacctccatt  2880
gagaatatca aagaattaaa ggccggctat atttcccaag tcgtccacaa gatctgcgag  2940
ctggtggaga aatatgacgc cgtgattgcg ctcgaagact taaattctgg gttcaagaac  3000
tcccgcgtga aggtggaaaa acaggtgtat cagaaattcg agaaaatgct gatcgacaaa  3060
ctcaattata tggtggataa gaagtccaac ccgtgtgcca caggggcgc gctgaagggc  3120
tatcagatca ccaacaagtt cgagagcttc aagagcatga gcacccagaa cgggtttatt  3180
ttctacatcc cggcgtggct cacctccaag attgacccga gcaccggctt cgtgaacctc  3240
ctgaagacaa agtatacctc cattgccgac agcaagaagt ttatctcctc cttcgaccgc  3300
attatgtatg tgccggagga ggacctcttc gagttcgccc tcgactacaa aaacttcagc  3360
cgcacagatg cggattacat caagaagtgg aagctgtact cctacgggaa caggatccgc  3420
atcttcagga atccaaaaaa aaataacgtc tttgactggg aggaagtgtg cctgacatcc  3480
gcctacaagg aactgttcaa taaatacggc atcaattacc agcagggcga cattcgcgcc  3540
ctcctctgtg agcagtccga caaagcgttt tactccagct tcatggccct catgtccctg  3600
atgctccaaa tgaggaatag catcacaggg cgcaccgacg tcgacttcct catcagcccg  3660
gtgaagaact ccgacgggat cttttacgac tcccgcaact atgaggcgca agagaatgcg  3720
atcctcccga agaacgccga tgcgaacggg gcctataata tcgccaggaa agtgctctgg  3780
gccatcgggc agttcaaaaa ggcggaggat gagaagctcg acaaggtgaa aattgccatt  3840
tccaacaagg agtggctgga gtacgcgcag acctccgtga agcactctgg tggttcaagc  3900
ggaggctcat ctggcagcga aactccgggc acttccgagt cagctactcc tgagtctagc  3960
ggcgggtcgt caggagggtc tgctcttgat tttctttcta ggcttccatt gccacctcca  4020
gtttctccaa tttgtacttt tgtgtcacct gctgcacaaa aggcattcca gcctccaagg  4080
tcatgcggat ctggcggctc aaaaagaacc gccgacggca gcgaattcga gcccaagaag  4140
aagaggaaag tctaa                                                    4155

SEQ ID NO: 187          moltype = AA  length = 845
FEATURE                 Location/Qualifiers
source                  1..845
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 187
MKRTADGSEF ESPKKKRKVT LNIEDEHRLH ETSKEPDVSL GSTWLSDFPQ AWAETGGMGL  60
AVRQAPLIIP LKATSTPVSI KQYPMSQEAR LGIKPHIQRL LDQGILVPCQ SPWNTPLLPV  120
KKPGTNDYRP VQDLREVNKR VEDIHPTVPN PYNLLSGLPP SHQWYTVLDL KDAFFCLRLH  180
PTSQPLFAFE WRDPEMGISG QLTWTRLPQG FKNSPTLFNE ALHRDLADFR IQHPDLILLQ  240
YVDDLLLAAT SELDCQQGTR ALLQTLGNLG YRASAKKAQI CQKQVKYLGY LLKEGQRWLT  300
EARKETVMGQ PTPKTPRQLR EFLGKAGFCR LFIPGFAEMA APLYPLTKPG TLFNWGPDQQ  360
KAYQEIKQAL LTAPALGLPD LTKPFELFVD EKQGYAKGVL TQKLGPWRRP VAYLSKKLDP  420
VAAGWPPCLR MVAAIAVLTK DAGKLTMGQP LVILAPHAVE ALVKQPPDRW LSNARMTHYQ  480
ALLLDTDRVQ FGPVVALNPA TLLPLPEEGL QHNCLDILAE AHGTRPDLTD QPLPDADHTW  540
YTDGSSLLQE GQRKAGAAVT TETEVIWAKA LPAGTSAQRA ELIALTQALK MAEGKKLNVY  600
TDSRYAFATA HIHGEIYRRR GWLTSEGKEI KNKDEILALL KALFLPKRLS IIHCPGHQKG  660
HSAEARGNRM ADQAARKAAI TETPDTSTLL IENSSPNSRL INSGGSSGGS SGSETPGTSE  720
SACLSYETEI LTVEYGLLPI GKIVEKRIEC TVYSVDNNGN IYTQPVAQWH DRGEQEVFEY  780
CLEDGSLIRA TKDHKFMTVD GQMLPIDEIF ERELDLMRVD NLPNSGGSKR TADGSEFEPK  840
KKRKV                                                           845

SEQ ID NO: 188          moltype = AA  length = 1384
FEATURE                 Location/Qualifiers
source                  1..1384
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 188
MKRTADGSEF ESPKKKRKVI KIATRKYLGK QNVYDIGVGE PHNFALKNGF IASNCATPES  60
SGGSSGGSSK LEKFTNCYSL SKTLRFKAIP VGKTQENIDN KRLLVEDEKR AEDYKGVKKL  120
LDRYYLSFIN DVLHSIKLKN LNNYISLFRK KTRTEKENKE LENLEINLRK EIAKAFKGNE  180
GYKSLFKKDI IETILPEFLD DKDEIALVNS FNGFTTAFTG FFDNRENMFS EEAKSTSIAF  240
RCINENLTRY ISNMDIFEKV DAIFDKHEVQ EIKEKILNSD YDVEDFFEGE FFNFVLTQEG  300
IDVYNAIIGG FVTESGEKIK GLNEYINLYN QKTKQKLPKF KPLYKQVLSD RESLSFYGEG  360
YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS AGIFVKNGPA ISTISKDIFG  420
```

```
                                   -continued
EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG SFSLEQLQEY ADADLSVVEK  480
LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV AIMKDLLDSV KSFENYIKAF  540
FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ KPYSKDKFKL YFQNPQFMGG  600
WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD VNGNYEKINY KLLPGPNKML  660
PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH KLIDFFKDSI SRYPKWSNAY  720
DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL VEEGKLYMFQ IYNKDFSDKS  780
HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK EELVVHPANS PIANKNPDNP  840
KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN TEVRVLLKHD DNPYVIGIDR  900
GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH SLLDKKEKER FEARQNWTSI  960
ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN SRVKVEKQVY QKFEKMLIDK  1020
LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI FYIPAWLTSK IDPSTGFVNL  1080
LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS RTDADYIKKW KLYSYGNRIR  1140
IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA LLCEQSDKAF YSSFMALMSL  1200
MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA ILPKNADANG AYNIARKVLW  1260
AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKHSGGSS GGSSGSETPG TSESATPESS  1320
GGSSGGSALD FLSRLPLPPP VSPICTFVSP AAQKAFQPPR SCGSGGSKRT ADGSEFEPKK  1380
KRKV                                                              1384

SEQ ID NO: 189          moltype = AA  length = 20
FEATURE                 Location/Qualifiers
source                  1..20
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 189
SGGSSGGSSG SETPGTSESA                                             20

SEQ ID NO: 190          moltype = AA  length = 12
FEATURE                 Location/Qualifiers
source                  1..12
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 190
TPESSGGSSG GS                                                     12

SEQ ID NO: 191          moltype = AA  length = 61
FEATURE                 Location/Qualifiers
source                  1..61
                        mol_type = protein
                        organism = synthetic construct
VARIANT                 4..60
                        note = Residues are present or absent
SEQUENCE: 191
GSSGSSGSSG SSGSSGSSGS SGSSGSSGSS GSSGSSGSSG SSGSSGSSGS SGSSGSSGSS  60
G                                                                 61

SEQ ID NO: 192          moltype = AA  length = 1300
FEATURE                 Location/Qualifiers
source                  1..1300
                        mol_type = protein
                        organism = Francisella tularensis
SEQUENCE: 192
MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF  60
FIEEILSSVC ISEDLLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK  120
NLFNQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANSDIT DIDEALEIIK SFKGWTTYFK  180
GFHENRKNVY SSNDIPTSII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE  240
ELTFDIDYKT SEVNQRVFSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI  300
NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA  360
AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY  420
ITQQIAPKNL DNPSKKEQEL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA  480
NFAAIPMIFD EIAQNKDNLA QISIKYQNQG KKDLLQASAE DDVKAIKDLL DQTNNLLHKL  540
KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF  600
ENSTLANGWD KNKEPDNTAI LFIKDDKYYL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK  660
LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GSPQKGYEKF EFNIEDCRKF  720
IDFYKQSISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ  780
GKLYLFQIYN KDFSAYSKGR PNLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKQSIPKK  840
ITHPAKEAIA NKNKDNPKKE SVFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI  900
NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI  960
EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKLVIEYN AIVVFEDLNF GFKRGRFKVE  1020
KQVYQKLEKM LIEKLNYLVF KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG  1080
FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG  1140
KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAAICGESD  1200
KKFFAKLTSV LNTILQMRNS KTGTELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY  1260
HIGLKGLMLL GRIKNNQEGK KLNLVIKNEE YFEFVQNRNN                       1300

SEQ ID NO: 193          moltype = DNA  length = 3681
FEATURE                 Location/Qualifiers
```

-continued

```
source                 1..3681
                       mol_type = genomic DNA
                       note = Lachnospiraceae bacterium
                       organism = unidentified
SEQUENCE: 193
tccaagctgg agaagtttac aaactgttac agcctctcca aaaccctcag gtttaaagcg  60
atcccggtgg gcaagaccca ggagaacatc gacaacaaga ggctcctggt ggaagacgag  120
aagcgcgccg aagactacaa gggcgtgaag aagctgctcg ataggtacta cctcagcttt  180
attaacgacg tgctgcacag catcaaactc aagaatctca acaactacat ctccctcttc  240
cgcaaaaaga cccgcaccga gaaggagaac aaggagctgg agaacctgga gatcaacctc  300
cgcaaggaaa tcgccaaagc gttcaagggc aatgaagggt acaagagcct cttcaagaaa  360
gacatcatcg aaactatcct cccagagttt ctcgatgaca aggacgagat cgcgctggtg  420
aactccttta acgggttcac aaccgcgttt accggcttct ttgataacag ggaaaatatg  480
ttctccgagg aggccaagtc caccagcatc gccttcaggt gtatcaacga gaacctcacc  540
cgctacattt ccaatatgga cattttcgag aaggtggatg cgatcttcga taagcacgag  600
gtgcaggaga tcaaagagaa gattctcaat tccgattatg acgtcgagga tttcttcgaa  660
ggggagttct ttaattttgt gctcacacaa gagggcattg acgtgtacaa cgcgattatc  720
gggggcttcg tcacagagtc cggggagaag attaaggggc tgaatgagta catcaatctg  780
tacaatcaga agaccaagca gaaactgccg aaattcaagc cgctctacaa gcaagtcctg  840
tccgataggg aaagcctctc cttctacggc gagggctata ccagcgacga ggaggtgctg  900
gaagtcttcc gcaacacact gaataagaat agcgagattt tctcctccat caagaagctc  960
gagaagctct ttaagaactt tgacgagtac agctccgccg ggattttcgt gaagaacggg  1020
ccggcgatca gcaccatctc caaggacatc tttggcgagt ggaacgtcat cagggacaag  1080
tggaacgccg agtacgacga catccacctg aagaagaagg cggtggtgac cgagaagtat  1140
gaggacgatc gcaggaagtc cttcaaaaaa atcggctcct tcagcctcga acagctccag  1200
gagtatgccg atgcggatct gtccgtcgtc gagaagctga aggaaatcat cattcagaag  1260
gtcgacgaga tctataaagt gtacgggtcc agcgagaagc tgttcgacgc cgactttgtg  1320
ctcgagaagt ccctcaaaaa gaatgacgcc gtggtggcca ttatgaaaga cctgctcgac  1380
tccgtgaagt ccttcgaaaa ttacattaaa gcgttctttg gggaggggaa ggaaactaac  1440
agggatgagt ccttctatgg cgactttgtc ctcgcgtacg acatcctgct gaaggtcgac  1500
cacatttacg acgcgatccg caactacgtg acacagaagc cgtactccaa agacaagttc  1560
aagctgtact tccagaaccc gcaatttatg gggggctggg acaaggataa agagacagac  1620
taccgcgcga caattctccg ctatggctcc aaatactatc tggccatcat ggacaagaag  1680
tacgcgaagt gcctgcagaa gatcgacaaa gacgacgtca atggcaacta tgaaaagatc  1740
aactacaagc tgctgccggg cccgaacaag atgctcccga aggtgttctt cagcaagaag  1800
tggatggcct actacaatcc aagcgaggat attcagaaaa tctataaaaa cgggaccttc  1860
aagaaggggg acatgtttaa cctcaacgac tgccacaagc tcattgattt cttcaaggat  1920
agcatttccc gctacccgaa atggtccaat gcgtacgatt ttaacttctc cgagacagaa  1980
aagtacaaag acatcgcggg cttttacagg gaggtggagg agcaagggta taaagtttct  2040
tttgaatccg cgagcaagaa ggaagtcgac aagctcgtcg aggagggcaa gctctacatg  2100
ttccaaattt ataacaagga cttttccgac aagagccatg ggaccccaaa cctccacacc  2160
atgtacttca aactgctctt tgacgagaac aaccacgggc aaatcaggct gagcggcggc  2220
gccgaattat tcatgcgcag ggcctccctc aagaaggaag agctggtcgt ccatccagcc  2280
aattccccga tcgcgaacaa gaacccggac aatccgaaaa agaccaccac cctgtcctac  2340
gacgtctaca aggacaaacg cttcagcgaa gaccagtacg aattacacat cccaattgcg  2400
attaataagt gcccaaagaa tatcttcaaa attaatacag aggtcagggt gctgctcaaa  2460
cacgacgaca atccgtatgt catcggcatt gacaggggcg agcgcaatct gctctatatc  2520
gtggtcgtgg atgggaaggg caatattgtg gagcagtact ccctgaacga gattatcaac  2580
aacttcaatg ggattaggat taagaccgac tatcacagcc tgctcgacaa gaaagaaaaa  2640
gagaggtttg aggcccgcca aaactggacc tccattgaga atatcaaaga attaaaggcc  2700
ggctatattt cccaagtcgt ccacaagatc tgcgagctgg tggagaaata tgacgccgtg  2760
attgcgctcg aagacttaaa ttctgggttc aagaactccc gcgtgaaggt ggaaaaacag  2820
gtgtatcaga aattcgagaa aatgctgatc gacaaactca attatatggt ggataagaag  2880
tccaacccgt gtgccacagg gggcgcgctg aagggctatc agatcaccaa caagttcgag  2940
agcttcaaga gcatgagcac ccagaacggg tttatttttct acatcccggc gtggctcacc  3000
tccaagattg acccgagcac cggcttcgtg aacctcctga agacaaagta tacctccatt  3060
gccgacagca agaagtttat ctcctccttc gaccgcatta tgtatgtgcc ggaggaggac  3120
ctcttcgagt tcgccctcga ctacaaaaac ttcagccgca cagatgcgga ttacatcaag  3180
aagtggaagc tgtactccta cgggaacagg atccgcatct tcaggaatcc aaaaaaaaat  3240
aacgtctttg actgggagga agtgtgcctg acatccgcct acaaggaact gttcaataaa  3300
tacggcatca attaccagca gggcgacatt cgcgccctcc tctgtgagca gtccgacaaa  3360
gcgttttact ccagcttcat ggccctcatg tccctgatgc tccaaatgag gaatagcatc  3420
acagggcgca ccgacgtcga cttcctcatc agcccggtga agaactccga cgggatcttt  3480
tacgactccc gcaactatga ggcgcaagag aatgcgatcc tcccgaagaa cgccgatgcg  3540
aacggggcct ataatatcgc caggaaagtg ctctgggcca tcgggcagtt caaaaaggcg  3600
gaggatgaga agctcgacaa ggtgaaaatt gccatttcca acaaggagtg gctggagtac  3660
gcgcagacct ccgtgaagca c                                            3681

SEQ ID NO: 194         moltype = AA  length = 1228
FEATURE                Location/Qualifiers
source                 1..1228
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 194
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
```

-continued

```
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  300
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  420
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  480
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  540
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  600
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  720
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS  780
YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY  840
IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK  900
AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK  960
KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS  1020
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  1080
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS  1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1200
AEDEKLDKVK IAISNKEWLE YAQTSVKH                                     1228

SEQ ID NO: 195          moltype = AA  length = 1228
FEATURE                 Location/Qualifiers
source                  1..1228
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 195
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGE KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV  300
LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD  360
KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ  420
KVDEIYKVYG SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET  480
NRDESFYGDF VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET  540
DYRATILRYG SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK  600
KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET  660
EKYKDIAGFY REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH  720
TMYFKLLFDE NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS  780
YDVYKDKRFS EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IDRGERNLLY  840
IVVVDGKGNI VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK  900
AGYISQVVHK ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK  960
KSNPCATGGA LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS  1020
IADSKKFISS FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK  1080
NNVFDWEEVC LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMANS  1140
ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK  1200
AEDEKLDKVK IAISNKEWLE YAQTSVKH                                     1228

SEQ ID NO: 196          moltype = AA  length = 1315
FEATURE                 Location/Qualifiers
source                  1..1315
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 196
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGIDGVK  300
LSSTEKKKMD EIAPPKGSVF TCPICEKRSI VGVTANLVHD HNHDTGWGRE WICDSCNTGL  360
GRFKDNPKFL EKVIEYLKKG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  420
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  480
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  540
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  600
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  660
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  720
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  780
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  840
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  900
TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  960
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1020
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1080
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1140
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1200
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1260
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH       1315
```

-continued

```
SEQ ID NO: 197         moltype = AA  length = 1501
FEATURE                Location/Qualifiers
source                 1..1501
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 197
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGAAPNI  300
HGWGKEGHEI ICKIAQTRLD ETAAKAVKEL LPESAEGDLS SLCLWADRVK FRYHWSSPLH  360
YINTPDACSY QYNRDCKDES GEKGRCVAGA IYNYTTQLLS YKTAASSQSQ YNLTEALLFV  420
SHFMGDIHQP LHVSYASDKG GNTIEVHWYT RKANLHHIWD SNIIETAEAD LYNSALEGMV  480
DALKKNITTE WADQVKRWET CTKKTACPDI YASEGIQAAC DWAYKGVTEG DTLEDEYFYS  540
RLPIVYQRLA QGGVRLAATL NRIFGGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN  600
FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK  660
SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK  720
KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI  780
RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL RYGSKYYLAI MDKKYAKCLQ  840
KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF  900
NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA GFYREVEEQG YKVSFESASK  960
KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL FDENNHGQIR LSGGAELFMR  1020
RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK RFSEDQYELH IPIAINKCPK  1080
NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK GNIVEQYSLN EIINNFNGIR  1140
IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV VHKICELVEK YDAVIALEDL  1200
NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT GGALKGYQIT NKFESFKSMS  1260
TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF ISSFDRIMYV PEEDLFEFAL  1320
DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ  1380
QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV DFLISPVKNS DGIFYDSRNY  1440
EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE WLEYAQTSVK  1500
H                                                                 1501

SEQ ID NO: 198         moltype = AA  length = 1454
FEATURE                Location/Qualifiers
source                 1..1454
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 198
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGARVRT  300
YPNVSHANTH YKNTVSSKLL PFTANYQLQL GELDNLNRAT FSHIQLQDRH ETKDVRTKIN  360
YDPVGWHNYQ FPYGDGSKSS WVMNRGHLVG YQFCGLNDEP RNLVAMTAWL NTGAYSGAND  420
SNPEGMLYYE NRLDSWLALH PDFWLDYKVT PIYSGNEVVP RQIELQYVGI DSSGELLTIR  480
LNSNKESIDE NGVTTVILEN SAPNINLDYL NGTATPKNGY TSDEEVLEVF RNTLNKNSEI  540
FSSIKKLEKL FKNFDEYSSA GIFVKNGPAI STISKDIFGE WNVIRDKWNA EYDDIHLKKK  600
AVVTEKYEDD RRKSFKKIGS FSLEQLQEYA DADLSVVEKL KEIIIQKVDE IYKVYGSSEK  660
LFDADFVLEK SLKKNDAVVA IMKDLLDSVK SFENYIKAFF GEGKETNRDE SFYGDFVLAY  720
DILLKVDHIY DAIRNYVTQK PYSKDKFKLY FQNPQFMGGW DKDKETDYRA TILRYGSKYY  780
LAIMDKKYAK CLQKIDKDDV NGNYEKINYK LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK  840
IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS RYPKWSNAYD FNFSETEKYK DIAGFYREVE  900
EQGYKVSFES ASKKEVDKLV EEGKLYMFQI YNKDFSDKSH GTPNLHTMYF KLLFDENNHG  960
QIRLSGGAEL FMRRASLKKE ELVVHPANSP IANKNPDNPK KTTTLSYDVY KDKRFSEDQY  1020
ELHIPIAINK CPKNIFKINT EVRVLLKHDD NPYVIGIARG ERNLLYIVVV DGKGNIVEQY  1080
SLNEIINNFN GIRIKTDYHS LLDKKEKERF EARQNWTSIE NIKELKAGYI SQVVHKICEL  1140
VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ KFEKMLIDKL NYMVDKKSNP CATGGALKGY  1200
QITNKFESFK SMSTQNGFIF YIPAWLTSKI DPSTGFVNLL KTKYTSIADS KKFISSFDRI  1260
MYVPEEDLFE FALDYKNFSR TDADYIKKWK LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA  1320
YKELFNKYGI NYQQGDIRAL LCEQSDKAFY SSFMALMSLM LQMRNSITGR TDVDFLISPV  1380
KNSDGIFYDS RNYEAQENAI LPKNADANGA YNIARKVLWA IGQFKKAEDE KLDKVKIAIS  1440
NKEWLEYAQT SVKH                                                   1454

SEQ ID NO: 199         moltype = AA  length = 1375
FEATURE                Location/Qualifiers
source                 1..1375
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 199
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
```

-continued

```
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGENQTT  300
GKGKNNSRPR YKSLEKAIKE FGSQILKEHP TDNQELRNNR LYLYYLQNGK DMYTGQDLDI  360
HNLSNYDIDH IVPQSFITDN SIDNLVLTSS AGNREKGDDV PPLEIVRKRK VFWEKLYQGN  420
LMSKRKFDYL TKAERGGLTG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH        1375

SEQ ID NO: 200          moltype = AA  length = 1375
FEATURE                 Location/Qualifiers
source                  1..1375
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 200
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGENQFT  300
NQGRRNSQQR LKGLTDSIKE FGSQILKEHP VENSQLQNDR LFLYYLQNGR DMYTGEELDI  360
DYLSQYDIDH IIPQAFIKDN SIDNRVLTSS KENRGKSDDV PSKDVVRKMK SYWSKLLSAK  420
LITQRKFDNL TKAERGGLTG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS  480
AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG  540
SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV  600
AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ  660
KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD  720
VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH  780
KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL  840
VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK  900
EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN  960
TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH  1020
SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN  1080
SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI  1140
FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS  1200
RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA  1260
LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA  1320
ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH        1375

SEQ ID NO: 201          moltype = AA  length = 1398
FEATURE                 Location/Qualifiers
source                  1..1398
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 201
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGVREGF  300
VTEKRAREMD GDMRRRAARN AKLFQEMQEK LNVQGKPSRA DLWRYQSVQR QNCQCAYCGS  360
PITFSNSEMD HIVPRAGQGS TNTRENLVAV CHRCNQSKGN TPFAIWAKNT SIEGVSVKEA  420
VERTRHWVTD TGMRSTDFKK FTKAVVERFQ RATMDEEIDA RSGYTSDEEV LEVFRNTLNK  480
NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH  540
LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG  600
SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF  660
VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG  720
SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE  780
DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY  840
REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE  900
NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS  960
EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI  1020
VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK  1080
ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA  1140
LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS  1200
```

-continued

```
FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC  1260
LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL  1320
ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK  1380
IAISNKEWLE YAQTSVKH                                                1398

SEQ ID NO: 202         moltype = AA  length = 1491
FEATURE                Location/Qualifiers
source                 1..1491
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 202
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGGKAQL  300
DIKNFPELYR TTERVYKKSG QSTKPVTVSN IHYSVLDGYG RSGEAYGIIT KDMIDMSAGY  360
REKWESKPEP SGWYSYFFKN TNQRATESDY KHSPKNVSKI SNNIKASILL SNGNVRNGYL  420
FDRSHLIADS LGGRPFRNNL ITGTRTQNVG NNDRKGGMQY IENKVLDHIK RNPKVHVYYK  480
ATPVYQGSEL LPRAVLVSAL SSDGFIDETV RVFNNVAGFN IDYQNGGLLS STADVDINNV  540
EENEIETTDD EIEEGGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF  600
VKNGPAISTI SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL  660
EQLQEYADAD LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK KNDAVVAIMK  720
DLLDSVKSFE NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI RNYVTQKPYS  780
KDKFKLYFQN PQFMGGWDKD KETDYRATIL RYGSKYYLAI MDKKYAKCLQ KIDKDDVNGN  840
YEKINYKLLP GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF NLNDCHKLID  900
FFKDSISRYP KWSNAYDFNF SETEKYKDIA GFYREVEEQG YKVSFESASK KEVDKLVEEG  960
KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL FDENNHGQIR LSGGAELFMR RASLKKEELV  1020
VHPANSPIAN KNPDNPKKTT TLSYDVYKDK RFSEDQYELH IPIAINKCPK NIFKINTEVR  1080
VLLKHDDNPY VIGIARGERN LLYIVVVDGK GNIVEQYSLN EIINNFNGIR IKTDYHSLLD  1140
KKEKERFEAR QNWTSIENIK ELKAGYISQV VHKICELVEK YDAVIALEDL NSGFKNSRVK  1200
VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT GGALKGYQIT NKFESFKSMS TQNGFIFYIP  1260
AWLTSKIDPS TGFVNLLKTK YTSIADSKKF ISSFDRIMYV PEEDLFEFAL DYKNFSRTDA  1320
DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ QGDIRALLCE  1380
QSDKAFYSSF MALMSLMLQM RNSITGRTDV DFLISPVKNS DGIFYDSRNY EAQENAILPK  1440
NADANGAYNI ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE WLEYAQTSVK H           1491

SEQ ID NO: 203         moltype = AA  length = 1351
FEATURE                Location/Qualifiers
source                 1..1351
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 203
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGSNNSY  300
LRAKVFETEH GVCQLCNVNA QELFLRLRDA PKSQRKNLLY ATWTSKLPLE QLNEMIRNPG  360
EGHFWQVDHI KPVYGGGGQC SLDNLQTLCT VCHKERTARQ AKERSQVRRQ SLASKGYTSD  420
EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI SKDIFGEWNV  480
IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAD LSVVEKLKEI  540
IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG  600
KETNRDESFY GDFVLAYDIL LKVDHIYDAI RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD  660
KETDYRATIL RYGSKYYLAI MDKKYAKCLQ KIDKDDVNGN YEKINYKLLP GPNKMLPKVF  720
FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF NLNDCHKLID FFKDSISRYP KWSNAYDFNF  780
SETEKYKDIA GFYREVEEQG YKVSFESASK KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP  840
NLHTMYFKLL FDENNHGQIR LSGGAELFMR RASLKKEELV VHPANSPIAN KNPDNPKKTT  900
TLSYDVYKDK RFSEDQYELH IPIAINKCPK NIFKINTEVR VLLKHDDNPY VIGIARGERN  960
LLYIVVVDGK GNIVEQYSLN EIINNFNGIR IKTDYHSLLD KKEKERFEAR QNWTSIENIK  1020
ELKAGYISQV VHKICELVEK YDAVIALEDL NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM  1080
VDKKSNPCAT GGALKGYQIT NKFESFKSMS TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK  1140
YTSIADSKKF ISSFDRIMYV PEEDLFEFAL DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN  1200
PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ QGDIRALLCE QSDKAFYSSF MALMSLMLQM  1260
RNSITGRTDV DFLISPVKNS DGIFYDSRNY EAQENAILPK NADANGAYNI ARKVLWAIGQ  1320
FKKAEDEKLD KVKIAISNKE WLEYAQTSVK H                                 1351

SEQ ID NO: 204         moltype = AA  length = 1367
FEATURE                Location/Qualifiers
source                 1..1367
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 204
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
```

-continued

```
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGDEPGV  300
ATGNGQPVTG NWLAGASQGD GVPIPSQIAD QLRGKEFKSW RDFREQFWMA VSKDPSALEN  360
LSPSNRYFVS QGLAPYAVPE EHLGSKEKFE IHHVVPLESG GALYNIDNLV IVTPKRHSEI  420
HKELKLKRKE KGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG  480
PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ  540
EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA VVAIMKDLLD  600
SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV TQKPYSKDKF  660
KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK DDVNGNYEKI  720
NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND CHKLIDFFKD  780
SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD KLVEEGKLYM  840
FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL KKEELVVHPA  900
NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK INTEVRVLLK  960
HDDNPYVIGI ARGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD YHSLLDKKEK  1020
ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF KNSRVKVEKQ  1080
VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG FIFYIPAWLT  1140
SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN FSRTDADYIK  1200
KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI RALLCEQSDK  1260
AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE NAILPKNADA  1320
NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH                 1367

SEQ ID NO: 205          moltype = AA  length = 1299
FEATURE                 Location/Qualifiers
source                  1..1299
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 205
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGRLREQ  300
ALMRDKGLCL HCKNNRKIKV ADMVDHIIPI KVDPSLKLKL ENLQSLCNPC HNRKTAEDKK  360
KYGGYTSDEE VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD EYSSAGIFVK NGPAISTISK  420
DIFGEWNVIR DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF KKIGSFSLEQ LQEYADADLS  480
VVEKLKEIII QKVDEIYKVY GSSEKLFDAD FVLEKSLKKN DAVVAIMKDL LDSVKSFENY  540
IKAFFGEGKE TNRDESFYGD FVLAYDILLK VDHIYDAIRN YVTQKPYSKD KFKLYFQNPQ  600
FMGGWDKDKE TDYRATILRY GSKYYLAIMD KKYAKCLQKI DKDDVNGNYE KINYKLLPGP  660
NKMLPKVFFS KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL NDCHKLIDFF KDSISRYPKW  720
SNAYDFNFSE TEKYKDIAGF YREVEEQGYK VSFESASKKE VDKLVEEGKL YMFQIYNKDF  780
SDKSHGTPNL HTMYFKLLFD ENNHGQIRLS GGAELFMRRA SLKKEELVVH PANSPIANKN  840
PDNPKKTTTL SYDVYKDKRF SEDQYELHIP IAINKCPKNI FKINTEVRVL LKHDDNPYVI  900
GIARGERNLL YIVVVDGKGN IVEQYSLNEI INNFNGIRIK TDYHSLLDKK EKERFEARQN  960
WTSIENIKEL KAGYISQVVH KICELVEKYD AVIALEDLNS GFKNSRVKVE KQVYQKFEKM  1020
LIDKLNYMVD KKSNPCATGG ALKGYQITNK FESFKSMSTQ NGFIFYIPAW LTSKIDPSTG  1080
FVNLLKTKYT SIADSKKFIS SFDRIMYVPE EDLFEFALDY KNFSRTDADY IKKWKLYSYG  1140
NRIRIFRNPK KNNVFDWEEV CLTSAYKELF NKYGINYQQG DIRALLCEQS DKAFYSSFMA  1200
LMSLMLQMRN SITGRTDVDF LISPVKNSDG IFYDSRNYEA QENAILPKNA DANGAYNIAR  1260
KVLWAIGQFK KAEDEKLDKV KIAISNKEWL EYAQTSVKH                          1299

SEQ ID NO: 206          moltype = AA  length = 1328
FEATURE                 Location/Qualifiers
source                  1..1328
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 206
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGMNYFI  300
VEVSEQEVKR EKEKARELRR SQWWKNRIAR GICHYCGEIF PPEELTMDHL VPVVRGGKST  360
RGNVVPACKE CNNRKKYLLP VEWEEYLDSL ESGYTSDEEV LEVFRNTLNK NSEIFSSIKK  420
LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK  480
YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF  540
VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV  600
DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK  660
KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT  720
FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV  780
SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG  840
GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI  900
AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII  960
NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA  1020
VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF  1080
ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE  1140
DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN  1200
```

-continued

```
KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI  1260
FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE  1320
YAQTSVKH                                                           1328

SEQ ID NO: 207          moltype = AA  length = 1441
FEATURE                 Location/Qualifiers
source                  1..1441
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 207
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGPPSSF  300
SKAKKEAVKI YLDYPTSFYC GCDITWKNKK KGIPELESCG YQVRKQEKRA SRIEWEHVVP  360
AWQFGHQRQC WQKGGRKNCT RNDKQFKSME ADLHNLVPAI GEVNGDRSNF RFSQWNGSKG  420
AFYGQCAFKV DFKGRVAEPP AQSRGAIART YLYMNNEYKF NLSKAQRQLM EAWNKQYPVS  480
TWECTRDERI AKIQGNHNQF VYKACGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN  540
FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK  600
SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK  660
KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI  720
RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL RYGSKYYLAI MDKKYAKCLQ  780
KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF  840
NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA GFYREVEEQG YKVSFESASK  900
KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL FDENNHGQIR LSGGAELFMR  960
RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK RFSEDQYELH IPIAINKCPK  1020
NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK GNIVEQYSLN EIINNFNGIR  1080
IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV VHKICELVEK YDAVIALEDL  1140
NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT GGALKGYQIT NKFESFKSMS  1200
TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF ISSFDRIMYV PEEDLFEFAL  1260
DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ  1320
QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV DFLISPVKNS DGIFYDSRNY  1380
EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE WLEYAQTSVK  1440
H                                                                  1441

SEQ ID NO: 208          moltype = AA  length = 1506
FEATURE                 Location/Qualifiers
source                  1..1506
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 208
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGWRRRR  300
EDARAAPGLL GRLPVLPVAA AAELPPVPGG PRGPGELAKY GLPGLAQLKS RESYVLCYDP  360
RTRGALWVVE QLRPERLRGD GDRRECDFRE DDSVHAYHRA TNADYRGSGF DRGHLAAAAN  420
HRWSQKAMDD TFYLSNVAPQ VPHLNQNAWN NLEKYSRSLT RSYQNVYVCT GPLFLPRTEA  480
DGKSYVKYQV IGKNHVAVPT HFFKVLILEA AGGQIELRTY VMPNAPVDEA IPLERFLVPI  540
ESIERASGLL FVPNILARAG SLKAITAGSK GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE  600
KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE  660
DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL  720
EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH  780
IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY  840
AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK  900
KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF  960
ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA  1020
ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI  1080
NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE QYSLNEIINN  1140
FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI  1200
ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES  1260
FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL  1320
FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY  1380
GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY  1440
DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA  1500
QTSVKH                                                             1506

SEQ ID NO: 209          moltype = AA  length = 1508
FEATURE                 Location/Qualifiers
source                  1..1508
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 209
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
```

-continued

```
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGWSKEG  300
HVMTCRIAQG LLNDEAAHAV KMLLPEYVNG DLSALCVWPD QVRHWYKYKW TSPLHFIDTP  360
DKACNFDYER DCHDQHGVKD MCVAGAIQNF TTQLSHYREG TSDRRYNMTE ALLFLSHFMG  420
DIHQPMHVGF TSDAGGNSID LRWFRHKSNL HHVWDREIIL TAAKDYYAKD INLLEEDIEG  480
NFTDGIWSDD LASWRECGNV FSCVNKFATE SINIACKWGY KGVEAGETLS DDYFNSRLPI  540
VMKRVAQGGI RLAMLLNNVF GASQQEDSVV ATGYTSDEEV LEVFRNTLNK NSEIFSSIKK  600
LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK  660
YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF  720
VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV  780
DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK  840
KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT  900
FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV  960
SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG  1020
GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI  1080
AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII  1140
NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA  1200
VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF  1260
ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE  1320
DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN  1380
KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI  1440
FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE  1500
YAQTSVKH                                                         1508

SEQ ID NO: 210          moltype = AA  length = 1318
FEATURE                 Location/Qualifiers
source                  1..1318
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 210
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGMKSKI  300
SEYTEKEFLE FVKDIYTNNK KKFPTEESHI QAVLEFKKLT EHPSGSDLLY YPNENREDSP  360
AGVVKEVKEW RASKGLPGFK AGGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE  420
YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK  480
KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND  540
AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY  600
VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID  660
KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN  720
DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV  780
DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS  840
LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF  900
KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT  960
DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG  1020
FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN  1080
GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK  1140
NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD  1200
IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ  1260
ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKH    1318

SEQ ID NO: 211          moltype = AA  length = 1318
FEATURE                 Location/Qualifiers
source                  1..1318
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 211
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGMELKN  300
SISDYTEAEF VQLLKEIEKE NVAATDDVLD VLLEHFVKIT EHPDGTDLIY YPSDNRDDSP  360
EGIVKEIKEW RAANGKPGFK QGGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE  420
YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK  480
KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND  540
AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY  600
VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID  660
KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN  720
DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV  780
DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS  840
LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF  900
```

-continued

```
KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT  960
DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG  1020
FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN  1080
GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK  1140
NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD  1200
IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ  1260
ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKH    1318

SEQ ID NO: 212          moltype = AA  length = 1341
FEATURE                 Location/Qualifiers
source                  1..1341
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 212
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGGRTPT  300
AEERRIANAL GALPCIACYM HGVISNEVSL HHIAGRTAPG CHKKQLPLCR WHHQHAAPAE  360
VREKYPWLVP VHADGVVGGK KEFTLLNKSE MELLADAYEM ANIMHGYTSD EEVLEVFRNT  420
LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD  480
DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK  540
VYGSSEKLFD ADFVLEKSLK KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY  600
GDFVLAYDIL LKVDHIYDAI RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL  660
RYGSKYYLAI MDKKYAKCLQ KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN  720
PSEDIQKIYK NGTFKKGDMF NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA  780
GFYREVEEQG YKVSFESASK KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL  840
FDENNHGQIR LSGGAELFMR RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK  900
RFSEDQYELH IPIAINKCPK NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK  960
GNIVEQYSLN EIINNFNGIR IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV  1020
VHKICELVEK YDAVIALEDL NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT  1080
GGALKGYQIT NKFESFKSMS TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF  1140
ISSFDRIMYV PEEDLFEFAL DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE  1200
EVCLTSAYKE LFNKYGINYQ QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV  1260
DFLISPVKNS DGIFYDSRNY EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD  1320
KVKIAISNKE WLEYAQTSVK H                                           1341

SEQ ID NO: 213          moltype = AA  length = 1336
FEATURE                 Location/Qualifiers
source                  1..1336
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 213
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGMEWKD  300
IKGYEGHYQV SNTGEVYSIK SGKTLKHQIP KDGYHRIGLF KGGKGKTFQV HRLVAIHFCE  360
GYEEGLVVDH KDGNKDNNLS TNLRWVTQKI NVENQMSRGT GYTSDEEVLE VFRNTLNKNS  420
EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK  480
KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS  540
EKLFDADFVL EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL  600
AYDILLKVDH IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK  660
YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI  720
QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE  780
VEEQGYKVSF ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN  840
HGQIRLSGGA ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED  900
QYELHIPIAI NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE  960
QYSLNEIINN FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC  1020
ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK  1080
GYQITNKFES FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD  1140
RIMYVPEEDL FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT  1200
SAYKELFNKY GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS  1260
PVKNSDGIFY DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA  1320
ISNKEWLEYA QTSVKH                                                 1336

SEQ ID NO: 214          moltype = AA  length = 1501
FEATURE                 Location/Qualifiers
source                  1..1501
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 214
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
```

-continued

```
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGWGALG  300
HATVAYVAQH YVSPEAASWA QGILGSSSSS YLASIASWAD EYRLTSAGKW SASLHFIDAE  360
DNPPTNCNVD YERDCGSSGC SISAIANYTQ RVSDSSLSSE NHAEALRFLV HFIGDMTQPL  420
HDEAYAVGGN KINVTFDGYH DNLHSDWDTY MPQKLIGGHA LSDAESWAKT LVQNIESGNY  480
TAQAIGWIKG DNISEPITTA TRWASDANAL VCTVVMPHGA AALQTGDLYP TYYDSVIDTI  540
ELQIAKGGYR LANWINEIHG SEIAKGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN  600
FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK  660
SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK  720
KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI  780
RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL RYGSKYYLAI MDKKYAKCLQ  840
KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF  900
NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA GFYREVEEQG YKVSFESASK  960
KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL FDENNHGQIR LSGGAELFMR  1020
RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK RFSEDQYELH IPIAINKCPK  1080
NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK GNIVEQYSLN EIINNFNGIR  1140
IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV VHKICELVEK YDAVIALEDL  1200
NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT GGALKGYQIT NKFESFKSMS  1260
TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF ISSFDRIMYV PEEDLFEFAL  1320
DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ  1380
QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV DFLISPVKNS DGIFYDSRNY  1440
EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE WLEYAQTSVK  1500
H                                                                  1501

SEQ ID NO: 215          moltype = AA  length = 1498
FEATURE                 Location/Qualifiers
source                  1..1498
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 215
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGWGNLG  300
HETVAYIAQS FVASSTESFC QNILGDDSTS YLANVATWAD TYKYTDAGEF SKPYHFIDAQ  360
DNPPQSCGVD YDRDCGSAGC SISAIQNYTN ILLESPNGSE ALNALKFVVH IIGDIHQPLH  420
DENLEAGGNG IDVTYDGETT NLHHIWDTNM PEEAAGGYSL SVAKTYADLL TERIKTGTYS  480
SKKDSWTDGI DIKDPVSTSM IWAADANTYV CSTVLDDGLA YINSTDLSGE YYDKSQPVFE  540
ELIAKAGYRL AAWLDLIASQ PSGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE  600
YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK  660
KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND  720
AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY  780
VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID  840
KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN  900
DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV  960
DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS  1020
LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF  1080
KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT  1140
DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG  1200
FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN  1260
GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK  1320
NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD  1380
IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ  1440
ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKH    1498

SEQ ID NO: 216          moltype = AA  length = 1317
FEATURE                 Location/Qualifiers
source                  1..1317
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 216
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGMELKH  300
SISDYTEAEF LQLVTTICNA DTSSEEELVK LVTHFEEMTE HPSGSDLIYY PKEGDDDSPS  360
GIVNTVKQWR AANGKSGFKQ GGYTSDEEVL EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY  420
SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK  480
IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV LEKSLKKNDA  540
VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD HIYDAIRNYV  600
TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK YAKCLQKIDK  660
DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF KKGDMFNLND  720
CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS FESASKKEVD  780
```

-continued

```
KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG AELFMRRASL  840
KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA INKCPKNIFK  900
INTEVRVLLK HDDNPYVIGI ARGERNLLYI VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD  960
YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV IALEDLNSGF  1020
KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE SFKSMSTQNG  1080
FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED LFEFALDYKN  1140
FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK YGINYQQGDI  1200
RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE  1260
NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY AQTSVKH     1317

SEQ ID NO: 217         moltype = AA  length = 1328
FEATURE                Location/Qualifiers
source                 1..1328
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 217
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGMLLTG  300
KLYKEEKQKF YDAQNGKCLI CQRELNPDVQ ANHLDHDHEL NGPKAGKVRG LLCNLCNAAE  360
GQMKHKFNRS GLKGQGVDYL EWLENLLTYL KSGYTSDEEV LEVFRNTLNK NSEIFSSIKK  420
LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK  480
YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF  540
VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV  600
DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK  660
KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT  720
FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV  780
SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG  840
GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI  900
AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII  960
NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA  1020
VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF  1080
ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE  1140
DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN  1200
KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI  1260
FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE  1320
YAQTSVKH                                                           1328

SEQ ID NO: 218         moltype = AA  length = 1438
FEATURE                Location/Qualifiers
source                 1..1438
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 218
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGSFSHA  300
KNEAVKIYRD HPVSFYCGCE IRWQGKKGIP DLESCGYQVR KNENRASRIE WEHVVPAWQF  360
GHQLQCWQQG GRKNCTRTSP EFNQMEADLH NLTPAIGEVN GDRSNFSFSQ WNGVDGVTYG  420
QCEMQVNFKE RTAMPPERAR GAIARTYLYM SEQYGLRLSK AQSQLMQAWN NQYPVSEWEC  480
VRDQRIEKVQ GNSNRFVREQ CPGYTSDEEV LEVFRNTLNK NSEIFSSIKK LEKLFKNFDE  540
YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH LKKKAVVTEK YEDDRRKSFK  600
KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG SSEKLFDADF VLEKSLKKND  660
AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF VLAYDILLKV DHIYDAIRNY  720
VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG SKYYLAIMDK KYAKCLQKID  780
KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE DIQKIYKNGT FKKGDMFNLN  840
DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY REVEEQGYKV SFESASKKEV  900
DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE NNHGQIRLSG GAELFMRRAS  960
LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS EDQYELHIPI AINKCPKNIF  1020
KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI VEQYSLNEII NNFNGIRIKT  1080
DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK ICELVEKYDA VIALEDLNSG  1140
FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA LKGYQITNKF ESFKSMSTQN  1200
GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS FDRIMYVPEE DLFEFALDYK  1260
NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC LTSAYKELFN KYGINYQQGD  1320
IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL ISPVKNSDGI FYDSRNYEAQ  1380
ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK IAISNKEWLE YAQTSVKH    1438

SEQ ID NO: 219         moltype = AA  length = 1480
FEATURE                Location/Qualifiers
source                 1..1480
                       mol_type = protein
                       organism = synthetic construct
```

-continued

```
SEQUENCE: 219
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGADLPA  300
LPGGPAGGTG ELAKYGLPGV AQLRSRESYV LSYDPRTRGA LWVLEQLRPE RLRGDGDRSA  360
CDFREDDSVH AYHRATNADY RGSGFDRGHL AAAANHRWSQ RAMDDTFYLS NVAPQVPHLN  420
QNAWNNLERY SRSLTRTYQN VYVCTGPLFL PRTEADGKSY VKYQVIGKNH VAVPTHFFKV  480
LILEAAGGQI ELRSYVMPNA PVDETIPLER FLVPIESIER ASGLLFVPNI LARAGNLKAI  540
TAGSGYTSDE EVLEVFRNTL NKNSEIFSSI KKLEKLFKNF DEYSSAGIFV KNGPAISTIS  600
KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT EKYEDDRRKS FKKIGSFSLE QLQEYADADL  660
SVVEKLKEII IQKVDEIYKV YGSSEKLFDA DFVLEKSLKK NDAVVAIMKD LLDSVKSFEN  720
YIKAFFGEGK ETNRDESFYG DFVLAYDILL KVDHIYDAIR NYVTQKPYSK DKFKLYFQNP  780
QFMGGWDKDK ETDYRATILR YGSKYYLAIM DKKYAKCLQK IDKDDVNGNY EKINYKLLPG  840
PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFN LNDCHKLIDF FKDSISRYPK  900
WSNAYDFNFS ETEKYKDIAG FYREVEEQGY KVSFESASKK EVDKLVEEGK LYMFQIYNKD  960
FSDKSHGTPN LHTMYFKLLF DENNHGQIRL SGGAELFMRR ASLKKEELVV HPANSPIANK  1020
NPDNPKKTTT LSYDVYKDKR FSEDQYELHI PIAINKCPKN IFKINTEVRV LLKHDDNPYV  1080
IGIARGERNL LYIVVVDGKG NIVEQYSLNE IINNFNGIRI KTDYHSLLDK KEKERFEARQ  1140
NWTSIENIKE LKAGYISQVV HKICELVEKY DAVIALEDLN SGFKNSRVKV EKQVYQKFEK  1200
MLIDKLNYMV DKKSNPCATG GALKGYQITN KFESFKSMST QNGFIFYIPA WLTSKIDPST  1260
GFVNLLKTKY TSIADSKKFI SSFDRIMYVP EEDLFEFALD YKNFSRTDAD YIKKWKLYSY  1320
GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL FNKYGINYQQ GDIRALLCEQ SDKAFYSSFM  1380
ALMSLMLQMR NSITGRTDVD FLISPVKNSD GIFYDSRNYE AQENAILPKN ADANGAYNIA  1440
RKVLWAIGQF KKAEDEKLDK VKIAISNKEW LEYAQTSVKH                        1480

SEQ ID NO: 220         moltype = AA  length = 1382
FEATURE                Location/Qualifiers
source                 1..1382
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 220
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGEKNSK  300
DAQKMINEMQ KRNRQTNERI EEIIRTTGKE NAKYLIEKIK LHDMQEGKCL YSLEAIPLED  360
LLNNPFNYEV DHIIPRSVSF DNSFNNKVLV KQEENSKKGN RTPFQYLSSS DSKISYETFK  420
KHILNLAKGK GRISKTKKEY LLEERDGYTS DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK  480
NFDEYSSAGI FVKNGPAIST ISKDIFGEWN VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR  540
KSFKKIGSFS LEQLQEYADA DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL  600
KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA  660
IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL  720
QKIDKDDVNG NYEKINYKLL PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM  780
FNLNDCHKLI DFFKDSISRY PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS  840
KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM  900
RRASLKKEEL VVHPANSPIA NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP  960
KNIFKINTEV RVLLKHDDNP YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI  1020
RIKTDYHSLL DKKEKERFEA RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED  1080
LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM  1140
STQNGFIFYI PAWLTSKIDP STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA  1200
LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY  1260
QQGDIRALLC EQSDKAFYSS FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN  1320
YEAQENAILP KNADANGAYN IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV  1380
KH                                                                 1382

SEQ ID NO: 221         moltype = AA  length = 1335
FEATURE                Location/Qualifiers
source                 1..1335
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 221
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGDGNAH  300
TVNPSKLVSH RLGDGLTVQQ IDRACTPALW CALTREKDFD EKNGLPARED RAIRVHGHEI  360
KSSDYIQVFS KRKKTDSDRD ETPFGAIAVR GGFVEIGPSG YTSDEEVLEV FRNTLNKNSE  420
IFSSIKKLEK LFKNFDEYSS AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK  480
KAVVTEKYED DRRKSFKKIG SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE  540
KLFDADFVLE KSLKKNDAVV AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA  600
YDILLKVDHI YDAIRNYVTQ KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY  660
YLAIMDKKYA KCLQKIDKDD VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ  720
```

-continued

```
KIYKNGTFKK GDMFNLNDCH KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV  780
EEQGYKVSFE SASKKEVDKL VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH  840
GQIRLSGGAE LFMRRASLKK EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ  900
YELHIPIAIN KCPKNIFKIN TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ  960
YSLNEIINNF NGIRIKTDYH SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE  1020
LVEKYDAVIA LEDLNSGFKN SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG  1080
YQITNKFESF KSMSTQNGFI FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR  1140
IMYVPEEDLF EFALDYKNFS RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS  1200
AYKELFNKYG INYQQGDIRA LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP  1260
VKNSDGIFYD SRNYEAQENA ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI  1320
SNKEWLEYAQ TSVKH                                                   1335

SEQ ID NO: 222          moltype = AA  length = 1381
FEATURE                 Location/Qualifiers
source                  1..1381
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 222
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGENQYT  300
NQGKSNSQQR LKRLEKSLKE LGSKILKENI PAKLSKIDNN ALQNDRLYLY YLQNGKDMYT  360
GDDLDIDRLS NYDIDHIIPQ AFLKDNSIDN KVLVSSASNR GKSDDFPSLE VVKKRKTFWY  420
QLLKSKLISQ RKFDNLTKAE RGGLLGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN  480
FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK  540
SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK  600
KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI  660
RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL RYGSKYYLAI MDKKYAKCLQ  720
KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF  780
NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA GFYREVEEQG YKVSFESASK  840
KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL FDENNHGQIR LSGGAELFMR  900
RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK RFSEDQYELH IPIAINKCPK  960
NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK GNIVEQYSLN EIINNFNGIR  1020
IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV VHKICELVEK YDAVIALEDL  1080
NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT GGALKGYQIT NKFESFKSMS  1140
TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF ISSFDRIMYV PEEDLFEFAL  1200
DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ  1260
QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV DFLISPVKNS DGIFYDSRNY  1320
EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE WLEYAQTSVK  1380
H                                                                  1381

SEQ ID NO: 223          moltype = AA  length = 1347
FEATURE                 Location/Qualifiers
source                  1..1347
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 223
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGFPNFV  300
GEPKSKDILK LRLYEQQHGK CLYSGKEINL GRLNEKGYVE IDHALPFSRT WDDSFNNKVL  360
VLGSENQNKG NQTPYEYFNG KDNSREWQEF KARVETSRFP RSKKQRILLQ KGYTSDEEVL  420
EVFRNTLNKN SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK  480
WNAEYDDIHL KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ EYADADLSVV EKLKEIIIQK  540
VDEIYKVYGS SEKLFDADFV LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN  600
RDESFYGDFV LAYDILLKVD HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD  660
YRATILRYGS KYYLAIMDKK YAKCLQKIDK DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK  720
WMAYYNPSED IQKIYKNGTF KKGDMFNLND CHKLIDFFKD SISRYPKWSN AYDFNFSETE  780
KYKDIAGFYR EVEEQGYKVS FESASKKEVD KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT  840
MYFKLLFDEN NHGQIRLSGG AELFMRRASL KKEELVVHPA NSPIANKNPD NPKKTTTLSY  900
DVYKDKRFSE DQYELHIPIA INKCPKNIFK INTEVRVLLK HDDNPYVIGI ARGERNLLYI  960
VVVDGKGNIV EQYSLNEIIN NFNGIRIKTD YHSLLDKKEK ERFEARQNWT SIENIKELKA  1020
GYISQVVHKI CELVEKYDAV IALEDLNSGF KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK  1080
SNPCATGGAL KGYQITNKFE SFKSMSTQNG FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI  1140
ADSKKFISSF DRIMYVPEED LFEFALDYKN FSRTDADYIK KWKLYSYGNR IRIFRNPKKN  1200
NVFDWEEVCL TSAYKELFNK YGINYQQGDI RALLCEQSDK AFYSSFMALM SLMLQMRNSI  1260
TGRTDVDFLI SPVKNSDGIF YDSRNYEAQE NAILPKNADA NGAYNIARKV LWAIGQFKKA  1320
EDEKLDKVKI AISNKEWLEY AQTSVKH                                      1347

SEQ ID NO: 224          moltype = AA  length = 1390
FEATURE                 Location/Qualifiers
source                  1..1390
```

-continued

```
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 224
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGSPENI  300
FKDKNNRIKE FAKGISAYSG ANLTDGDFDG AKEELDHIIP RSHKKYGTLN DEANLICVTR  360
GDNKNKGNRI FCLRDLADNY KLKQFETTDD LEIEKKIADT IWDANKKDFK FGNYRSFINL  420
TPQEQKAFRH ALFLADENPI KQAVIRAINN RNRTGYTSDE EVLEVFRNTL NKNSEIFSSI  480
KKLEKLFKNF DEYSSAGIFV KNGPAISTIS KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT  540
EKYEDDRRKS FKKIGSFSLE QLQEYADADL SVVEKLKEII IQKVDEIYKV YGSSEKLFDA  600
DFVLEKSLKK NDAVVAIMKD LLDSVKSFEN YIKAFFGEGK ETNRDESFYG DFVLAYDILL  660
KVDHIYDAIR NYVTQKPYSK DKFKLYFQNP QFMGGWDKDK ETDYRATILR YGSKYYLAIM  720
DKKYAKCLQK IDKDDVNGNY EKINYKLLPG PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN  780
GTFKKGDMFN LNDCHKLIDF FKDSISRYPK WSNAYDFNFS ETEKYKDIAG FYREVEEQGY  840
KVSFESASKK EVDKLVEEGK LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF DENNHGQIRL  900
SGGAELFMRR ASLKKEELVV HPANSPIANK NPDNPKKTTT LSYDVYKDKR FSEDQYELHI  960
PIAINKCPKN IFKINTEVRV LLKHDDNPYV IGIARGERNL LYIVVVDGKG NIVEQYSLNE  1020
IINNFNGIRI KTDYHSLLDK KEKERFEARQ NWTSIENIKE LKAGYISQVV HKICELVEKY  1080
DAVIALEDLN SGFKNSRVKV EKQVYQKFEK MLIDKLNYMV DKKSNPCATG GALKGYQITN  1140
KFESFKSMST QNGFIFYIPA WLTSKIDPST GFVNLLKTKY TSIADSKKFI SSFDRIMYVP  1200
EEDLFEFALD YKNFSRTDAD YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL  1260
FNKYGINYQQ GDIRALLCEQ SDKAFYSSFM ALMSLMLQMR NSITGRTDVD FLISPVKNSD  1320
GIFYDSRNYE AQENAILPKN ADANGAYNIA RKVLWAIGQF KKAEDEKLDK VKIAISNKEW  1380
LEYAQTSVKH                                                        1390

SEQ ID NO: 225          moltype = AA  length = 1337
FEATURE                 Location/Qualifiers
source                  1..1337
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 225
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGHKHRC  300
QVCGIVIEVD VGPYSQGAHI RPLGRKHGGP DVESNMLCLC PNDHVRFDNG ALYITDDLKV  360
VNALNGEVIG PLRVHPRHVI DLDHIRYHRS QLPNIPLEGS SGYTSDEEVL EVFRNTLNKN  420
SEIFSSIKKL EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL  480
KKKAVVTEKY EDDRRKSFKK IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS  540
SEKLFDADFV LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV  600
LAYDILLKVD HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS  660
KYYLAIMDKK YAKCLQKIDK DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED  720
IQKIYKNGTF KKGDMFNLND CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR  780
EVEEQGYKVS FESASKKEVD KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN  840
NHGQIRLSGG AELFMRRASL KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE  900
DQYELHIPIA INKCPKNIFK INTEVRVLLK HDDNPYVIGI ARGERNLLYI VVVDGKGNIV  960
EQYSLNEIIN NFNGIRIKTD YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI  1020
CELVEKYDAV IALEDLNSGF KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL  1080
KGYQITNKFE SFKSMSTQNG FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF  1140
DRIMYVPEED LFEFALDYKN FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL  1200
TSAYKELFNK YGINYQQGDI RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI  1260
SPVKNSDGIF YDSRNYEAQE NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI  1320
AISNKEWLEY AQTSVKH                                                 1337

SEQ ID NO: 226          moltype = AA  length = 1393
FEATURE                 Location/Qualifiers
source                  1..1393
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 226
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGSEAEE  300
VIFGFADTTE ITTTADAIEC QISNDASIDE TEKHQLVKSR RGQGIFRSRL EQVESRCRVT  360
GVQLKNHLIA SHIKPWAVSN NQERLDGHNG LLLAPHVDHL FDKGFISFED NGEMIVSEKL  420
NLDVLKAWSI SQGNYGYFSK QQQEYMCYHR ENVFKKLGYT SDEEVLEVFR NTLNKNSEIF  480
SSIKKLEKLF KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA  540
VVTEKYEDDR RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL  600
FDADFVLEKS LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD  660
ILLKVDHIYD AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL  720
```

-continued

```
AIMDKKYAKC LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI  780
YKNGTFKKGD MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE  840
QGYKVSFESA SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ  900
IRLSGGAELF MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE  960
LHIPIAINKC PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS  1020
LNEIINNFNG IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV  1080
EKYDAVIALE DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ  1140
ITNKFESFKS MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM  1200
YVPEEDLFEF ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY  1260
KELFNKYGIN YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK  1320
NSDGIFYDSR NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN  1380
KEWLEYAQTS VKH                                                     1393

SEQ ID NO: 227         moltype = AA  length = 1356
FEATURE                Location/Qualifiers
source                 1..1356
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 227
SKLEKFTNCY SLSKTLRFKA IPVGKTQENI DNKRLLVEDE KRAEDYKGVK KLLDRYYLSF  60
INDVLHSIKL KNLNNYISLF RKKTRTEKEN KELENLEINL RKEIAKAFKG NEGYKSLFKK  120
DIIETILPEF LDDKDEIALV NSFNGFTTAF TGFFDNRENM FSEEAKSTSI AFRCINENLT  180
RYISNMDIFE KVDAIFDKHE VQEIKEKILN SDYDVEDFFE GEFFNFVLTQ EGIDVYNAII  240
GGFVTESGEK IKGLNEYINL YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGSSGDEPGV  300
ATGNGQPVTG NWLAGASQGD GVPIPSQIAD QLRGKEFKSW RDFREQFWVA VANDPELVKY  360
FRKTNAKGMR DGLSPFTPKA EQAGGRDKYE IHHVVQISQG GAVYDIDNLR VMTPKMHIQV  420
GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP AISTISKDIF  480
GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE YADADLSVVE  540
KLKEIIIQKV DEIYKVYGSS EKLFDADFVL EKSLKKNDAV VAIMKDLLDS VKSFENYIKA  600
FFGEGKETNR DESFYGDFVL AYDILLKVDH IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG  660
GWDKDKETDY RATILRYGSK YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM  720
LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA  780
YDFNFSETEK YKDIAGFYRE VEEQGYKVSF ESASKKEVDK LVEEGKLYMF QIYNKDFSDK  840
SHGTPNLHTM YFKLLFDENN HGQIRLSGGA ELFMRRASLK KEELVVHPAN SPIANKNPDN  900
PKKTTTLSYD VYKDKRFSED QYELHIPIAI NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA  960
RGERNLLYIV VVDGKGNIVE QYSLNEIINN FNGIRIKTDY HSLLDKKEKE RFEARQNWTS  1020
IENIKELKAG YISQVVHKIC ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV YQKFEKMLID  1080
KLNYMVDKKS NPCATGGALK GYQITNKFES FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN  1140
LLKTKYTSIA DSKKFISSFD RIMYVPEEDL FEFALDYKNF SRTDADYIKK WKLYSYGNRI  1200
RIFRNPKKNN VFDWEEVCLT SAYKELFNKY GINYQQGDIR ALLCEQSDKA FYSSFMALMS  1260
LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY DSRNYEAQEN AILPKNADAN GAYNIARKVL  1320
WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA QTSVKH                            1356

SEQ ID NO: 228         moltype = AA  length = 1395
FEATURE                Location/Qualifiers
source                 1..1395
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 228
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGSPENIFK DKNNRIKEFA KGISAYSGAN LTDGDFDGAK EELDHIIPRS  300
HKKYGTLNDE ANLICVTRGD NKNKGNRIFC LRDLADNYKL KQFETTDDLE IEKKIADTIW  360
DANKKDFKFG NYRSFINLTP QEQKAFRHAL FLADENPIKQ AVIRAINNRN RTGSSGEKIK  420
GLNEYINLYN QKTKQKLPKF KPLYKQVLSD RESLSFYGEG YTSDEEVLEV FRNTLNKNSE  480
IFSSIKKLEK LFKNFDEYSS AGIFVKNGPA ISTISKDIFG EWNVIRDKWN AEYDDIHLKK  540
KAVVTEKYED DRRKSFKKIG SFSLEQLQEY ADADLSVVEK LKEIIIQKVD EIYKVYGSSE  600
KLFDADFVLE KSLKKNDAVV AIMKDLLDSV KSFENYIKAF FGEGKETNRD ESFYGDFVLA  660
YDILLKVDHI YDAIRNYVTQ KPYSKDKFKL YFQNPQFMGG WDKDKETDYR ATILRYGSKY  720
YLAIMDKKYA KCLQKIDKDD VNGNYEKINY KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ  780
KIYKNGTFKK GDMFNLNDCH KLIDFFKDSI SRYPKWSNAY DFNFSETEKY KDIAGFYREV  840
EEQGYKVSFE SASKKEVDKL VEEGKLYMFQ IYNKDFSDKS HGTPNLHTMY FKLLFDENNH  900
GQIRLSGGAE LFMRRASLKK EELVVHPANS PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ  960
YELHIPIAIN KCPKNIFKIN TEVRVLLKHD DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ  1020
YSLNEIINNF NGIRIKTDYH SLLDKKEKER FEARQNWTSI ENIKELKAGY ISQVVHKICE  1080
LVEKYDAVIA LEDLNSGFKN SRVKVEKQVY QKFEKMLIDK LNYMVDKKSN PCATGGALKG  1140
YQITNKFESF KSMSTQNGFI FYIPAWLTSK IDPSTGFVNL LKTKYTSIAD SKKFISSFDR  1200
IMYVPEEDLF EFALDYKNFS RTDADYIKKW KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS  1260
AYKELFNKYG INYQQGDIRA LLCEQSDKAF YSSFMALMSL MLQMRNSITG RTDVDFLISP  1320
VKNSDGIFYD SRNYEAQENA ILPKNADANG AYNIARKVLW AIGQFKKAED EKLDKVKIAI  1380
SNKEWLEYAQ TSVKH                                                   1395

SEQ ID NO: 229         moltype = AA  length = 1361
FEATURE                Location/Qualifiers
```

-continued

```
source                  1..1361
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 229
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGDEPGVAT GNGQPVTGNW LAGASQGDGV PIPSQIADQL RGKEFKSWRD  300
FREQFWVAVA NDPELVKYFR KTNAKGMRDG LSPFTPKAEQ AGGRDKYEIH HVVQISQGGA  360
VYDIDNLRVM TPKMHIQVGS SGEKIKGLNE YINLYNQKTK QKLPKFKPLY KQVLSDRESL  420
SFYGEGYTSD EEVLEVFRNT LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI  480
SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAD  540
LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD ADFVLEKSLK KNDAVVAIMK DLLDSVKSFE  600
NYIKAFFGEG KETNRDESFY GDFVLAYDIL LKVDHIYDAI RNYVTQKPYS KDKFKLYFQN  660
PQFMGGWDKD KETDYRATIL RYGSKYYLAI MDKKYAKCLQ KIDKDDVNGN YEKINYKLLP  720
GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK NGTFKKGDMF NLNDCHKLID FFKDSISRYP  780
KWSNAYDFNF SETEKYKDIA GFYREVEEQG YKVSFESASK KEVDKLVEEG KLYMFQIYNK  840
DFSDKSHGTP NLHTMYFKLL FDENNHGQIR LSGGAELFMR RASLKKEELV VHPANSPIAN  900
KNPDNPKKTT TLSYDVYKDK RFSEDQYELH IPIAINKCPK NIFKINTEVR VLLKHDDNPY  960
VIGIARGERN LLYIVVVDGK GNIVEQYSLN EIINNFNGIR IKTDYHSLLD KKEKERFEAR  1020
QNWTSIENIK ELKAGYISQV VHKICELVEK YDAVIALEDL NSGFKNSRVK VEKQVYQKFE  1080
KMLIDKLNYM VDKKSNPCAT GGALKGYQIT NKFESFKSMS TQNGFIFYIP AWLTSKIDPS  1140
TGFVNLLKTK YTSIADSKKF ISSFDRIMYV PEEDLFEFAL DYKNFSRTDA DYIKKWKLYS  1200
YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE LFNKYGINYQ QGDIRALLCE QSDKAFYSSF  1260
MALMSLMLQM RNSITGRTDV DFLISPVKNS DGIFYDSRNY EAQENAILPK NADANGAYNI  1320
ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE WLEYAQTSVK H                      1361

SEQ ID NO: 230          moltype = AA  length = 1342
FEATURE                 Location/Qualifiers
source                  1..1342
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 230
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGHKHRCQV CGIVIEVDVG PYSQGAHIRP LGRKHGGPDV ESNMLCLCPN  300
DHVRFDNGAL YITDDLKVVN ALNGEVIGPL RVHPRHVIDL DHIRYHRSQL PNIPLEGSSG  360
SSGEKIKGLN EYINLYNQKT KQKLPKFKPL YKQVLSDRES LSFYGEGYTS DEEVLEVFRN  420
TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI FVKNGPAIST ISKDIFGEWN VIRDKWNAEY  480
DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS LEQLQEYADA DLSVVEKLKE IIIQKVDEIY  540
KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE GKETNRDESF  600
YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK DKETDYRATI  660
LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL PGPNKMLPKV FFSKKWMAYY  720
NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY PKWSNAYDFN FSETEKYKDI  780
AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL  840
LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA NKNPDNPKKT TTLSYDVYKD  900
KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP YVIGIARGER NLLYIVVVDG  960
KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA RQNWTSIENI KELKAGYISQ  1020
VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA  1080
TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP STGFVNLLKT KYTSIADSKK  1140
FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW  1200
EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS FMALMSLMLQ MRNSITGRTD  1260
VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN IARKVLWAIG QFKKAEDEKL  1320
DKVKIAISNK EWLEYAQTSV KH                                           1342

SEQ ID NO: 231          moltype = AA  length = 1398
FEATURE                 Location/Qualifiers
source                  1..1398
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 231
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGSEAEEVI FGFADTTEIT TTADAIECQI SNDASIDETE KHQLVKSRRG  300
QGIFRSRLEQ VESRCRVTGV QLKNHLIASH IKPWAVSNNQ ERLDGHNGLL LAPHVDHLFD  360
KGFISFEDNG EMIVSEKLNL DVLKAWSISQ GNYGYFSKQQ QEYMCYHREN VFKKLGSSGE  420
KIKGLNEYIN LYNQKTKQKL PKFKPLYKQV LSDRESLSFY GEGYTSDEEV LEVFRNTLNK  480
NSEIFSSIKK LEKLFKNFDE YSSAGIFVKN GPAISTISKD IFGEWNVIRD KWNAEYDDIH  540
LKKKAVVTEK YEDDRRKSFK KIGSFSLEQL QEYADADLSV VEKLKEIIIQ KVDEIYKVYG  600
SSEKLFDADF VLEKSLKKND AVVAIMKDLL DSVKSFENYI KAFFGEGKET NRDESFYGDF  660
VLAYDILLKV DHIYDAIRNY VTQKPYSKDK FKLYFQNPQF MGGWDKDKET DYRATILRYG  720
```

-continued

```
SKYYLAIMDK KYAKCLQKID KDDVNGNYEK INYKLLPGPN KMLPKVFFSK KWMAYYNPSE  780
DIQKIYKNGT FKKGDMFNLN DCHKLIDFFK DSISRYPKWS NAYDFNFSET EKYKDIAGFY  840
REVEEQGYKV SFESASKKEV DKLVEEGKLY MFQIYNKDFS DKSHGTPNLH TMYFKLLFDE  900
NNHGQIRLSG GAELFMRRAS LKKEELVVHP ANSPIANKNP DNPKKTTTLS YDVYKDKRFS  960
EDQYELHIPI AINKCPKNIF KINTEVRVLL KHDDNPYVIG IARGERNLLY IVVVDGKGNI  1020
VEQYSLNEII NNFNGIRIKT DYHSLLDKKE KERFEARQNW TSIENIKELK AGYISQVVHK  1080
ICELVEKYDA VIALEDLNSG FKNSRVKVEK QVYQKFEKML IDKLNYMVDK KSNPCATGGA  1140
LKGYQITNKF ESFKSMSTQN GFIFYIPAWL TSKIDPSTGF VNLLKTKYTS IADSKKFISS  1200
FDRIMYVPEE DLFEFALDYK NFSRTDADYI KKWKLYSYGN RIRIFRNPKK NNVFDWEEVC  1260
LTSAYKELFN KYGINYQQGD IRALLCEQSD KAFYSSFMAL MSLMLQMRNS ITGRTDVDFL  1320
ISPVKNSDGI FYDSRNYEAQ ENAILPKNAD ANGAYNIARK VLWAIGQFKK AEDEKLDKVK  1380
IAISNKEWLE YAQTSVKH                                                1398

SEQ ID NO: 232         moltype = AA  length = 1352
FEATURE                Location/Qualifiers
source                 1..1352
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 232
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGFPNFVGE PKSKDILKLR LYEQQHGKCL YSGKEINLGR LNEKGYVEID  300
HALPFSRTWD DSFNNKVLVL GSENQNKGNQ TPYEYFNGKD NSREWQEFKA RVETSRFPRS  360
KKQRILLQKG SSGEKIKGLN EYINLYNQKT KQKLPKFKPL YKQVLSDRES LSFYGEGYTS  420
DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI FVKNGPAIST ISKDIFGEWN  480
VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS LEQLQEYADA DLSVVEKLKE  540
IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE  600
GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK  660
DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG NYEKINYKLL PGPNKMLPKV  720
FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI DFFKDSISRY PKWSNAYDFN  780
FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT  840
PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL VVHPANSPIA NKNPDNPKKT  900
TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV RVLLKHDDNP YVIGIARGER  960
NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL DKKEKERFEA RQNWTSIENI  1020
KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY  1080
MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI PAWLTSKIDP STGFVNLLKT  1140
KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR  1200
NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC EQSDKAFYSS FMALMSLMLQ  1260
MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP KNADANGAYN IARKVLWAIG  1320
QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KH                                1352

SEQ ID NO: 233         moltype = AA  length = 1386
FEATURE                Location/Qualifiers
source                 1..1386
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 233
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGENQYTNQ GKSNSQQRLK RLEKSLKELG SKILKENIPA KLSKIDNNAL  300
QNDRLYLYYL QNGKDMYTGD DLDIDRLSNY DIDHIIPQAF LKDNSIDNKV LVSSASNRGK  360
SDDFPSLEVV KKRKTFWYQL LKSKLISQRK FDNLTKAERG GLLGSSGEKI KGLNEYINLY  420
NQKTKQKLPK FKPLYKQVLS DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE  480
KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE  540
DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL  600
EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH  660
IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY  720
AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK  780
KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF  840
ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA  900
ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI  960
NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE QYSLNEIINN  1020
FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI  1080
ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES  1140
FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL  1200
FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY  1260
GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY  1320
DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA  1380
QTSVKH                                                             1386

SEQ ID NO: 234         moltype = AA  length = 1340
FEATURE                Location/Qualifiers
```

-continued

```
source                1..1340
                      mol_type = protein
                      organism = synthetic construct
SEQUENCE: 234
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGDGNAHTV NPSKLVSHRL GDGLTVQQID RACTPALWCA LTREKDFDEK  300
NGLPAREDRA IRVHGHEIKS SDYIQVFSKR KKTDSDRDET PFGAIAVRGG FVEIGPSGSS  360
GEKIKGLNEY INLYNQKTKQ KLPKFKPLYK QVLSDRESLS FYGEGYTSDE EVLEVFRNTL  420
NKNSEIFSSI KKLEKLFKNF DEYSSAGIFV KNGPAISTIS KDIFGEWNVI RDKWNAEYDD  480
IHLKKKAVVT EKYEDDRRKS FKKIGSFSLE QLQEYADADL SVVEKLKEII IQKVDEIYKV  540
YGSSEKLFDA DFVLEKSLKK NDAVVAIMKD LLDSVKSFEN YIKAFFGEGK ETNRDESFYG  600
DFVLAYDILL KVDHIYDAIR NYVTQKPYSK DKFKLYFQNP QFMGGWDKDK ETDYRATILR  660
YGSKYYLAIM DKKYAKCLQK IDKDDVNGNY EKINYKLLPG PNKMLPKVFF SKKWMAYYNP  720
SEDIQKIYKN GTFKKGDMFN LNDCHKLIDF FKDSISRYPK WSNAYDFNFS ETEKYKDIAG  780
FYREVEEQGY KVSFESASKK EVDKLVEEGK LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF  840
DENNHGQIRL SGGAELFMRR ASLKKEELVV HPANSPIANK NPDNPKKTTT LSYDVYKDKR  900
FSEDQYELHI PIAINKCPKN IFKINTEVRV LLKHDDNPYV IGIARGERNL LYIVVVDGKG  960
NIVEQYSLNE IINNFNGIRI KTDYHSLLDK KEKERFEARQ NWTSIENIKE LKAGYISQVV  1020
HKICELVEKY DAVIALEDLN SGFKNSRVKV EKQVYQKFEK MLIDKLNYMV DKKSNPCATG  1080
GALKGYQITN KFESFKSMST QNGFIFYIPA WLTSKIDPST GFVNLLKTKY TSIADSKKFI  1140
SSFDRIMYVP EEDLFEFALD YKNFSRTDAD YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE  1200
VCLTSAYKEL FNKYGINYQQ GDIRALLCEQ SDKAFYSSFM ALMSLMLQMR NSITGRTDVD  1260
FLISPVKNSD GIFYDSRNYE AQENAILPKN ADANGAYNIA RKVLWAIGQF KKAEDEKLDK  1320
VKIAISNKEW LEYAQTSVKH                                          1340

SEQ ID NO: 235        moltype = AA  length = 1387
FEATURE               Location/Qualifiers
source                1..1387
                      mol_type = protein
                      organism = synthetic construct
SEQUENCE: 235
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGEKNSKDA QKMINEMQKR NRQTNERIEE IIRTTGKENA KYLIEKIKLH  300
DMQEGKCLYS LEAIPLEDLL NNPFNYEVDH IIPRSVSFDN SFNNKVLVKQ EENSKKGNRT  360
PFQYLSSSDS KISYETFKKH ILNLAKGKGR ISKTKKEYLL EERDGSSGEK IKGLNEYINL  420
YNQKTKQKLP KFKPLYKQVL SDRESLSFYG EGYTSDEEVL EVFRNTLNKN SEIFSSIKKL  480
EKLFKNFDEY SSAGIFVKNG PAISTISKDI FGEWNVIRDK WNAEYDDIHL KKKAVVTEKY  540
EDDRRKSFKK IGSFSLEQLQ EYADADLSVV EKLKEIIIQK VDEIYKVYGS SEKLFDADFV  600
LEKSLKKNDA VVAIMKDLLD SVKSFENYIK AFFGEGKETN RDESFYGDFV LAYDILLKVD  660
HIYDAIRNYV TQKPYSKDKF KLYFQNPQFM GGWDKDKETD YRATILRYGS KYYLAIMDKK  720
YAKCLQKIDK DDVNGNYEKI NYKLLPGPNK MLPKVFFSKK WMAYYNPSED IQKIYKNGTF  780
KKGDMFNLND CHKLIDFFKD SISRYPKWSN AYDFNFSETE KYKDIAGFYR EVEEQGYKVS  840
FESASKKEVD KLVEEGKLYM FQIYNKDFSD KSHGTPNLHT MYFKLLFDEN NHGQIRLSGG  900
AELFMRRASL KKEELVVHPA NSPIANKNPD NPKKTTTLSY DVYKDKRFSE DQYELHIPIA  960
INKCPKNIFK INTEVRVLLK HDDNPYVIGI ARGERNLLYI VVVDGKGNIV EQYSLNEIIN  1020
NFNGIRIKTD YHSLLDKKEK ERFEARQNWT SIENIKELKA GYISQVVHKI CELVEKYDAV  1080
IALEDLNSGF KNSRVKVEKQ VYQKFEKMLI DKLNYMVDKK SNPCATGGAL KGYQITNKFE  1140
SFKSMSTQNG FIFYIPAWLT SKIDPSTGFV NLLKTKYTSI ADSKKFISSF DRIMYVPEED  1200
LFEFALDYKN FSRTDADYIK KWKLYSYGNR IRIFRNPKKN NVFDWEEVCL TSAYKELFNK  1260
YGINYQQGDI RALLCEQSDK AFYSSFMALM SLMLQMRNSI TGRTDVDFLI SPVKNSDGIF  1320
YDSRNYEAQE NAILPKNADA NGAYNIARKV LWAIGQFKKA EDEKLDKVKI AISNKEWLEY  1380
AQTSVKH                                                        1387

SEQ ID NO: 236        moltype = AA  length = 1485
FEATURE               Location/Qualifiers
source                1..1485
                      mol_type = protein
                      organism = synthetic construct
SEQUENCE: 236
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGADLPALP GGPAGGTGEL AKYGLPGVAQ LRSRESYVLS YDPRTRGALW  300
VLEQLRPERL RGDGDRSACD FREDDSVHAY HRATNADYRG SGFDRGHLAA AANHRWSQRA  360
MDDTFYLSNV APQVPHLNQN AWNNLERYSR SLTRTYQNVY VCTGPLFLPR TEADGKSYVK  420
YQVIGKNHVA VPTHFFKVLI LEAAGGQIEL RSYVMPNAPV DETIPLERFL VPIESIERAS  480
GLLFVPNILA RAGNLKAITA GSGSSGEKIK GLNEYINLYN QKTKQKLPKF KPLYKQVLSD  540
RESLSFYGEG YTSDEEVLEV FRNTLNKNSE IFSSIKKLEK LFKNFDEYSS AGIFVKNGPA  600
ISTISKDIFG EWNVIRDKWN AEYDDIHLKK KAVVTEKYED DRRKSFKKIG SFSLEQLQEY  660
```

-continued

```
ADADLSVVEK LKEIIIQKVD EIYKVYGSSE KLFDADFVLE KSLKKNDAVV AIMKDLLDSV  720
KSFENYIKAF FGEGKETNRD ESFYGDFVLA YDILLKVDHI YDAIRNYVTQ KPYSKDKFKL  780
YFQNPQFMGG WDKDKETDYR ATILRYGSKY YLAIMDKKYA KCLQKIDKDD VNGNYEKINY  840
KLLPGPNKML PKVFFSKKWM AYYNPSEDIQ KIYKNGTFKK GDMFNLNDCH KLIDFFKDSI  900
SRYPKWSNAY DFNFSETEKY KDIAGFYREV EEQGYKVSFE SASKKEVDKL VEEGKLYMFQ  960
IYNKDFSDKS HGTPNLHTMY FKLLFDENNH GQIRLSGGAE LFMRRASLKK EELVVHPANS  1020
PIANKNPDNP KKTTTLSYDV YKDKRFSEDQ YELHIPIAIN KCPKNIFKIN TEVRVLLKHD  1080
DNPYVIGIAR GERNLLYIVV VDGKGNIVEQ YSLNEIINNF NGIRIKTDYH SLLDKKEKER  1140
FEARQNWTSI ENIKELKAGY ISQVVHKICE LVEKYDAVIA LEDLNSGFKN SRVKVEKQVY  1200
QKFEKMLIDK LNYMVDKKSN PCATGGALKG YQITNKFESF KSMSTQNGFI FYIPAWLTSK  1260
IDPSTGFVNL LKTKYTSIAD SKKFISSFDR IMYVPEEDLF EFALDYKNFS RTDADYIKKW  1320
KLYSYGNRIR IFRNPKKNNV FDWEEVCLTS AYKELFNKYG INYQQGDIRA LLCEQSDKAF  1380
YSSFMALMSL MLQMRNSITG RTDVDFLISP VKNSDGIFYD SRNYEAQENA ILPKNADANG  1440
AYNIARKVLW AIGQFKKAED EKLDKVKIAI SNKEWLEYAQ TSVKH                  1485

SEQ ID NO: 237         moltype = AA  length = 1443
FEATURE                Location/Qualifiers
source                 1..1443
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 237
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGSFSHAKN EAVKIYRDHP VSFYCGCEIR WQGKKGIPDL ESCGYQVRKN  300
ENRASRIEWE HVVPAWQFGH QLQCWQQGGR KNCTRTSPEF NQMEADLHNL TPAIGEVNGD  360
RSNFSFSQWN GVDGVTYGQC EMQVNFKERT AMPPERARGA IARTYLYMSE QYGLRLSKAQ  420
SQLMQAWNNQ YPVSEWECVR DQRIEKVQGN SNRFVREQCP GSSGEKIKGL NEYINLYNQK  480
TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF  540
KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR  600
RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS  660
LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD  720
AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC  780
LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD  840
MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA  900
SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF  960
MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC  1020
PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG  1080
IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE  1140
DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS  1200
MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF  1260
ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN  1320
YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR  1380
NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS  1440
VKH                                                                1443

SEQ ID NO: 238         moltype = AA  length = 1333
FEATURE                Location/Qualifiers
source                 1..1333
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 238
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGMLLTGKL YKEEKQKFYD AQNGKCLICQ RELNPDVQAN HLDHDHELNG  300
PKAGKVRGLL CNLCNAAEGQ MKHKFNRSGL KGQGVDYLEW LENLLTYLKS GSSGEKIKGL  360
NEYINLYNQK TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF  420
SSIKKLEKLF KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA  480
VVTEKYEDDR RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL  540
FDADFVLEKS LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD  600
ILLKVDHIYD AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL  660
AIMDKKYAKC LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI  720
YKNGTFKKGD MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE  780
QGYKVSFESA SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ  840
IRLSGGAELF MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE  900
LHIPIAINKC PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS  960
LNEIINNFNG IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV  1020
EKYDAVIALE DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ  1080
ITNKFESFKS MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM  1140
YVPEEDLFEF ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY  1200
KELFNKYGIN YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK  1260
NSDGIFYDSR NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN  1320
KEWLEYAQTS VKH                                                     1333
```

-continued

```
SEQ ID NO: 239          moltype = AA  length = 1322
FEATURE                 Location/Qualifiers
source                  1..1322
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 239
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGMELKHSI SDYTEAEFLQ LVTTICNADT SSEEELVKLV THFEEMTEHP  300
SGSDLIYYPK EGDDDSPSGI VNTVKQWRAA NGKSGFKQGG SSGEKIKGLN EYINLYNQKT  360
KQKLPKFKPL YKQVLSDRES LSFYGEGYTS DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK  420
NFDEYSSAGI FVKNGPAIST ISKDIFGEWN VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR  480
KSFKKIGSFS LEQLQEYADA DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL  540
KKNDAVVAIM KDLLDSVKSF ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA  600
IRNYVTQKPY SKDKFKLYFQ NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL  660
QKIDKDDVNG NYEKINYKLL PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM  720
FNLNDCHKLI DFFKDSISRY PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS  780
KKEVDKLVEE GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM  840
RRASLKKEEL VVHPANSPIA NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP  900
KNIFKINTEV RVLLKHDDNP YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI  960
RIKTDYHSLL DKKEKERFEA RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED  1020
LNSGFKNSRV KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM  1080
STQNGFIFYI PAWLTSKIDP STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA  1140
LDYKNFSRTD ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY  1200
QQGDIRALLC EQSDKAFYSS FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN  1260
YEAQENAILP KNADANGAYN IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV  1320
KH                                                                1322

SEQ ID NO: 240          moltype = AA  length = 1503
FEATURE                 Location/Qualifiers
source                  1..1503
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 240
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGWGNLGHE TVAYIAQSFV ASSTESFCQN ILGDDSTSYL ANVATWADTY  300
KYTDAGEFSK PYHFIDAQDN PPQSCGVDYD RDCGSAGCSI SAIQNYTNIL LESPNGSEAL  360
NALKFVVHII GDIHQPLHDE NLEAGGNGID VTYDGETTNL HHIWDTNMPE EAAGGYSLSV  420
AKTYADLLTE RIKTGTYSSK KDSWTDGIDI KDPVSTSMIW AADANTYVCS TVLDDGLAYI  480
NSTDLSGEYY DKSQPVFEEL IAKAGYRLAA WLDLIASQPS GSSGEKIKGL NEYINLYNQK  540
TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF  600
KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR  660
RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS  720
LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD  780
AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC  840
LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD  900
MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA  960
SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF  1020
MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC  1080
PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG  1140
IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE  1200
DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS  1260
MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF  1320
ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN  1380
YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR  1440
NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS  1500
VKH                                                               1503

SEQ ID NO: 241          moltype = AA  length = 1506
FEATURE                 Location/Qualifiers
source                  1..1506
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 241
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGWGALGHA TVAYVAQHYV SPEAASWAQG ILGSSSSSYL ASIASWADEY  300
RLTSAGKWSA SLHFIDAEDN PPTNCNVDYE RDCGSSGCSI SAIANYTQRV SDSSLSSENH  360
```

-continued

```
AEALRFLVHF IGDMTQPLHD EAYAVGGNKI NVTFDGYHDN LHSDWDTYMP QKLIGGHALS  420
DAESWAKTLV QNIESGNYTA QAIGWIKGDN ISEPITTATR WASDANALVC TVVMPHGAAA  480
LQTGDLYPTY YDSVIDTIEL QIAKGGYRLA NWINEIHGSE IAKGSSGEKI KGLNEYINLY  540
NQKTKQKLPK FKPLYKQVLS DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE  600
KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE  660
DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL  720
EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH  780
IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY  840
AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK  900
KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF  960
ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA  1020
ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI  1080
NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE QYSLNEIINN  1140
FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI  1200
ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES  1260
FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL  1320
FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY  1380
GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY  1440
DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA  1500
QTSVKH                                                             1506

SEQ ID NO: 242         moltype = AA  length = 1341
FEATURE                Location/Qualifiers
source                 1..1341
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 242
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGMEWKDIK GYEGHYQVSN TGEVYSIKSG KTLKHQIPKD GYHRIGLFKG  300
GKGKTFQVHR LVAIHFCEGY EEGLVVDHKD GNKDNNLSTN LRWVTQKINV ENQMSRGTGS  360
SGEKIKGLNE YINLYNQKTK QKLPKFKPLY KQVLSDRESL SFYGEGYTSD EEVLEVFRNT  420
LNKNSEIFSS IKKLEKLFKN FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD  480
DIHLKKKAVV TEKYEDDRRK SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK  540
VYGSSEKLFD ADFVLEKSLK KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY  600
GDFVLAYDIL LKVDHIYDAI RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL  660
RYGSKYYLAI MDKKYAKCLQ KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN  720
PSEDIQKIYK NGTFKKGDMF NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA  780
GFYREVEEQG YKVSFESASK KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL  840
FDENNHGQIR LSGGAELFMR RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK  900
RFSEDQYELH IPIAINKCPK NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK  960
GNIVEQYSLN EIINNFNGIR IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV  1020
VHKICELVEK YDAVIALEDL NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT  1080
GGALKGYQIT NKFESFKSMS TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF  1140
ISSFDRIMYV PEEDLFEFAL DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE  1200
EVCLTSAYKE LFNKYGINYQ QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV  1260
DFLISPVKNS DGIFYDSRNY EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD  1320
KVKIAISNKE WLEYAQTSVK H                                            1341

SEQ ID NO: 243         moltype = AA  length = 1346
FEATURE                Location/Qualifiers
source                 1..1346
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 243
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGGRTPTAE ERRIANALGA LPCIACYMHG VISNEVSLHH IAGRTAPGCH  300
KKQLPLCRWH HQHAAPAEVR EKYPWLVPVH ADGVVGGKKE FTLLNKSEME LLADAYEMAN  360
IMHGSSGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS DRESLSFYGE GYTSDEEVLE  420
VFRNTLNKNS EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW  480
NAEYDDIHLK KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV  540
DEIYKVYGSS EKLFDADFVL EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR  600
DESFYGDFVL AYDILLKVDH IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY  660
RATILRYGSK YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW  720
MAYYNPSEDI QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK  780
YKDIAGFYRE VEEQGYKVSF ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM  840
YFKLLFDENN HGQIRLSGGA ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD  900
VYKDKRFSED QYELHIPIAI NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV  960
VVDGKGNIVE QYSLNEIINN FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG  1020
YISQVVHKIC ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS  1080
NPCATGGALK GYQITNKFES FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA  1140
```

-continued

```
DSKKFISSFD RIMYVPEEDL FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN  1200
VFDWEEVCLT SAYKELFNKY GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT  1260
GRTDVDFLIS PVKNSDGIFY DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE  1320
DEKLDKVKIA ISNKEWLEYA QTSVKH                                       1346

SEQ ID NO: 244          moltype = AA  length = 1513
FEATURE                 Location/Qualifiers
source                  1..1513
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 244
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGWSKEGHV MTCRIAQGLL NDEAAHAVKM LLPEYVNGDL SALCVWPDQV  300
RHWYKYKWTS PLHFIDTPDK ACNFDYERDC HDQHGVKDMC VAGAIQNFTT QLSHYREGTS  360
DRRYNMTEAL LFLSHFMGDI HQPMHVGFTS DAGGNSIDLR WFRHKSNLHH VWDREIILTA  420
AKDYYAKDIN LLEEDIEGNF TDGIWSDDLA SWRECGNVFS CVNKFATESI NIACKWGYKG  480
VEAGETLSDD YFNSRLPIVM KRVAQGGIRL AMLLNNVFGA SQQEDSVVAT GSSGEKIKGL  540
NEYINLYNQK TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF  600
SSIKKLEKLF KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA  660
VVTEKYEDDR RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL  720
FDADFVLEKS LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD  780
ILLKVDHIYD AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL  840
AIMDKKYAKC LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI  900
YKNGTFKKGD MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE  960
QGYKVSFESA SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ  1020
IRLSGGAELF MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE  1080
LHIPIAINKC PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS  1140
LNEIINNFNG IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV  1200
EKYDAVIALE DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ  1260
ITNKFESFKS MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM  1320
YVPEEDLFEF ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY  1380
KELFNKYGIN YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK  1440
NSDGIFYDSR NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN  1500
KEWLEYAQTS VKH                                                     1513

SEQ ID NO: 245          moltype = AA  length = 1446
FEATURE                 Location/Qualifiers
source                  1..1446
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 245
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGPPSSFSK AKKEAVKIYL DYPTSFYCGC DITWKNKKKG IPELESCGYQ  300
VRKQEKRASR IEWEHVVPAW QFGHQRQCWQ KGGRKNCTRN DKQFKSMEAD LHNLVPAIGE  360
VNGDRSNFRF SQWNGSKGAF YGQCAFKVDF KGRVAEPPAQ SRGAIARTYL YMNNEYKFNL  420
SKAQRQLMEA WNKQYPVSTW ECTRDERIAK IQGNHNQFVY KACGSSGEKI KGLNEYINLY  480
NQKTKQKLPK FKPLYKQVLS DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE  540
KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE  600
DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL  660
EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH  720
IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY  780
AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK  840
KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF  900
ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA  960
ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI  1020
NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE QYSLNEIINN  1080
FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI  1140
ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES  1200
FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL  1260
FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY  1320
GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY  1380
DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA  1440
QTSVKH                                                             1446

SEQ ID NO: 246          moltype = AA  length = 1304
FEATURE                 Location/Qualifiers
source                  1..1304
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 246
```

-continued

```
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGRLREQAL MRDKGLCLHC KNNRKIKVAD MVDHIIPIKV DPSLKLKLEN  300
LQSLCNPCHN RKTAEDKKKY GGSSGEKIKG LNEYINLYNQ KTKQKLPKFK PLYKQVLSDR  360
ESLSFYGEGY TSDEEVLEVF RNTLNKNSEI FSSIKKLEKL FKNFDEYSSA GIFVKNGPAI  420
STISKDIFGE WNVIRDKWNA EYDDIHLKKK AVVTEKYEDD RRKSFKKIGS FSLEQLQEYA  480
DADLSVVEKL KEIIIQKVDE IYKVYGSSEK LFDADFVLEK SLKKNDAVVA IMKDLLDSVK  540
SFENYIKAFF GEGKETNRDE SFYGDFVLAY DILLKVDHIY DAIRNYVTQK PYSKDKFKLY  600
FQNPQFMGGW DKDKETDYRA TILRYGSKYY LAIMDKKYAK CLQKIDKDDV NGNYEKINYK  660
LLPGPNKMLP KVFFSKKWMA YYNPSEDIQK IYKNGTFKKG DMFNLNDCHK LIDFFKDSIS  720
RYPKWSNAYD FNFSETEKYK DIAGFYREVE EQGYKVSFES ASKKEVDKLV EEGKLYMFQI  780
YNKDFSDKSH GTPNLHTMYF KLLFDENNHG QIRLSGGAEL FMRRASLKKE ELVVHPANSP  840
IANKNPDNPK KTTTLSYDVY KDKRFSEDQY ELHIPIAINK CPKNIFKINT EVRVLLKHDD  900
NPYVIGIARG ERNLLYIVVV DGKGNIVEQY SLNEIINNFN GIRIKTDYHS LLDKKEKERF  960
EARQNWTSIE NIKELKAGYI SQVVHKICEL VEKYDAVIAL EDLNSGFKNS RVKVEKQVYQ  1020
KFEKMLIDKL NYMVDKKSNP CATGGALKGY QITNKFESFK SMSTQNGFIF YIPAWLTSKI  1080
DPSTGFVNLL KTKYTSIADS KKFISSFDRI MYVPEEDLFE FALDYKNFSR TDADYIKKWK  1140
LYSYGNRIRI FRNPKKNNVF DWEEVCLTSA YKELFNKYGI NYQQGDIRAL LCEQSDKAFY  1200
SSFMALMSLM LQMRNSITGR TDVDFLISPV KNSDGIFYDS RNYEAQENAI LPKNADANGA  1260
YNIARKVLWA IGQFKKAEDE KLDKVKIAIS NKEWLEYAQT SVKH                   1304

SEQ ID NO: 247         moltype = AA  length = 1356
FEATURE                Location/Qualifiers
source                 1..1356
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 247
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGSNNSYLR AKVFETEHGV CQLCNVNAQE LFLRLRDAPK SQRKNLLYAT  300
WTSKLPLEQL NEMIRNPGEG HFWQVDHIKP VYGGGGQCSL DNLQTLCTVC HKERTARQAK  360
ERSQVRRQSL ASKGSSGEKI KGLNEYINLY NQKTKQKLPK FKPLYKQVLS DRESLSFYGE  420
GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE KLFKNFDEYS SAGIFVKNGP AISTISKDIF  480
GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE DDRRKSFKKI GSFSLEQLQE YADADLSVVE  540
KLKEIIIQKV DEIYKVYGSS EKLFDADFVL EKSLKKNDAV VAIMKDLLDS VKSFENYIKA  600
FFGEGKETNR DESFYGDFVL AYDILLKVDH IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG  660
GWDKDKETDY RATILRYGSK YYLAIMDKKY AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM  720
LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA  780
YDFNFSETEK YKDIAGFYRE VEEQGYKVSF ESASKKEVDK LVEEGKLYMF QIYNKDFSDK  840
SHGTPNLHTM YFKLLFDENN HGQIRLSGGA ELFMRRASLK KEELVVHPAN SPIANKNPDN  900
PKKTTTLSYD VYKDKRFSED QYELHIPIAI NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA  960
RGERNLLYIV VVDGKGNIVE QYSLNEIINN FNGIRIKTDY HSLLDKKEKE RFEARQNWTS  1020
IENIKELKAG YISQVVHKIC ELVEKYDAVI ALEDLNSGFK NSRVKVEKQV YQKFEKMLID  1080
KLNYMVDKKS NPCATGGALK GYQITNKFES FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN  1140
LLKTKYTSIA DSKKFISSFD RIMYVPEEDL FEFALDYKNF SRTDADYIKK WKLYSYGNRI  1200
RIFRNPKKNN VFDWEEVCLT SAYKELFNKY GINYQQGDIR ALLCEQSDKA FYSSFMALMS  1260
LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY DSRNYEAQEN AILPKNADAN GAYNIARKVL  1320
WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA QTSVKH                            1356

SEQ ID NO: 248         moltype = AA  length = 1403
FEATURE                Location/Qualifiers
source                 1..1403
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 248
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGVREGFVT EKRAREMDGD MRRRAARNAK LFQEMQEKLN VQGKPSRADL  300
WRYQSVQRQN CQCAYCGSPI TFSNSEMDHI VPRAGQGSTN TRENLVAVCH RCNQSKGNTP  360
FAIWAKNTSI EGVSVKEAVE RTRHWVTDTG MRSTDFKKFT KAVVERFQRA TMDEEIDARS  420
GSSGEKIKGL NEYINLYNQK TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR  480
NTLNKNSEIF SSIKKLEKLF KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE  540
YDDIHLKKKA VVTEKYEDDR RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI  600
YKVYGSSEKL FDADFVLEKS LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES  660
FYGDFVLAYD ILLKVDHIYD AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT  720
ILRYGSKYYL AIMDKKYAKC LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY  780
YNPSEDIQKI YKNGTFKKGD MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD  840
IAGFYREVEE QGYKVSFESA SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK  900
LLFDENNHGQ IRLSGGAELF MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK  960
DKRFSEDQYE LHIPIAINKC PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD  1020
```

-continued

```
GKGNIVEQYS LNEIINNFNG IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS  1080
QVVHKICELV EKYDAVIALE DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC  1140
ATGGALKGYQ ITNKFESFKS MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK  1200
KFISSFDRIM YVPEEDLFEF ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD  1260
WEEVCLTSAY KELFNKYGIN YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT  1320
DVDFLISPVK NSDGIFYDSR NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK  1380
LDKVKIAISN KEWLEYAQTS VKH                                          1403

SEQ ID NO: 249         moltype = AA  length = 1380
FEATURE                Location/Qualifiers
source                 1..1380
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 249
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGENQFTNQ GRRNSQQRLK GLTDSIKEFG SQILKEHPVE NSQLQNDRLF  300
LYYLQNGRDM YTGEELDIDY LSQYDIDHII PQAFIKDNSI DNRVLTSSKE NRGKSDDVPS  360
KDVVRKMKSY WSKLLSAKLI TQRKFDNLTK AERGGLTGSS GEKIKGLNEY INLYNQKTKQ  420
KLPKFKPLYK QVLSDRESLS FYGEGYTSDE EVLEVFRNTL NKNSEIFSSI KKLEKLFKNF  480
DEYSSAGIFV KNGPAISTIS KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT EKYEDDRRKS  540
FKKIGSFSLE QLQEYADADL SVVEKLKEII IQKVDEIYKV YGSSEKLFDA DFVLEKSLKK  600
NDAVVAIMKD LLDSVKSFEN YIKAFFGEGK ETNRDESFYG DFVLAYDILL KVDHIYDAIR  660
NYVTQKPYSK DKFKLYFQNP QFMGGWDKDK ETDYRATILR YGSKYYLAIM DKKYAKCLQK  720
IDKDDVNGNY EKINYKLLPG PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFN  780
LNDCHKLIDF FKDSISRYPK WSNAYDFNFS ETEKYKDIAG FYREVEEQGY KVSFESASKK  840
EVDKLVEEGK LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF DENNHGQIRL SGGAELFMRR  900
ASLKKEELVV HPANSPIANK NPDNPKKTTT LSYDVYKDKR FSEDQYELHI PIAINKCPKN  960
IFKINTEVRV LLKHDDNPYV IGIARGERNL LYIVVVDGKG NIVEQYSLNE IINNFNGIRI  1020
KTDYHSLLDK KEKERFEARQ NWTSIENIKE LKAGYISQVV HKICELVEKY DAVIALEDLN  1080
SGFKNSRVKV EKQVYQKFEK MLIDKLNYMV DKKSNPCATG GALKGYQITN KFESFKSMST  1140
QNGFIFYIPA WLTSKIDPST GFVNLLKTKY TSIADSKKFI SSFDRIMYVP EEDLFEFALD  1200
YKNFSRTDAD YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL FNKYGINYQQ  1260
GDIRALLCEQ SDKAFYSSFM ALMSLMLQMR NSITGRTDVD FLISPVKNSD GIFYDSRNYE  1320
AQENAILPKN ADANGAYNIA RKVLWAIGQF KKAEDEKLDK VKIAISNKEW LEYAQTSVKH  1380

SEQ ID NO: 250         moltype = AA  length = 1459
FEATURE                Location/Qualifiers
source                 1..1459
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 250
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGARVRTYP NVSHANTHYK NTVSSKLLPF TANYQLQLGE LDNLNRATFS  300
HIQLQDRHET KDVRTKINYD PVGWHNYQFP YGDGSKSSWV MNRGHLVGYQ FCGLNDEPRN  360
LVAMTAWLNT GAYSGANDSN PEGMLYYENR LDSWLALHPD FWLDYKVTPI YSGNEVVPRQ  420
IELQYVGIDS SGELLTIRLN SNKESIDENG VTTVILENSA PNINLDYLNG TATPKNGSSG  480
EKIKGLNEYI NLYNQKTKQK LPKFKPLYKQ VLSDRESLSF YGEGYTSDEE VLEVFRNTLN  540
KNSEIFSSIK KLEKLFKNFD EYSSAGIFVK NGPAISTISK DIFGEWNVIR DKWNAEYDDI  600
HLKKKAVVTE KYEDDRRKSF KKIGSFSLEQ LQEYADADLS VVEKLKEIII QKVDEIYKVY  660
GSSEKLFDAD FVLEKSLKKN DAVVAIMKDL LDSVKSFENY IKAFFGEGKE TNRDESFYGD  720
FVLAYDILLK VDHIYDAIRN YVTQKPYSKD KFKLYFQNPQ FMGGWDKDKE TDYRATILRY  780
GSKYYLAIMD KKYAKCLQKI DKDDVNGNYE KINYKLLPGP NKMLPKVFFS KKWMAYYNPS  840
EDIQKIYKNG TFKKGDMFNL NDCHKLIDFF KDSISRYPKW SNAYDFNFSE TEKYKDIAGF  900
YREVEEQGYK VSFESASKKE VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL HTMYFKLLFD  960
ENNHGQIRLS GGAELFMRRA SLKKEELVVH PANSPIANKN PDNPKKTTTL SYDVYKDKRF  1020
SEDQYELHIP IAINKCPKNI FKINTEVRVL LKHDDNPYVI GIARGERNLL YIVVVDGKGN  1080
IVEQYSLNEI INNFNGIRIK TDYHSLLDKK EKERFEARQN WTSIENIKEL KAGYISQVVH  1140
KICELVEKYD AVIALEDLNS GFKNSRVKVE KQVYQKFEKM LIDKLNYMVD KKSNPCATGG  1200
ALKGYQITNK FESFKSMSTQ NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT SIADSKKFIS  1260
SFDRIMYVPE EDLFEFALDY KNFSRTDADY IKKWKLYSYG NRIRIFRNPK KNNVFDWEEV  1320
CLTSAYKELF NKYGINYQQG DIRALLCEQS DKAFYSSFMA LMSLMLQMRN SITGRTDVDF  1380
LISPVKNSDG IFYDSRNYEA QENAILPKNA DANGAYNIAR KVLWAIGQFK KAEDEKLDKV  1440
KIAISNKEWL EYAQTSVKH                                               1459

SEQ ID NO: 251         moltype = AA  length = 1506
FEATURE                Location/Qualifiers
source                 1..1506
                       mol_type = protein
                       organism = synthetic construct
SEQUENCE: 251
```

-continued

```
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGAAPNIHG WGKEGHEIIC KIAQTRLDET AAKAVKELLP ESAEGDLSSL  300
CLWADRVKFR YHWSSPLHYI NTPDACSYQY NRDCKDESGE KGRCVAGAIY NYTTQLLSYK  360
TAASSQSQYN LTEALLFVSH FMGDIHQPLH VSYASDKGGN TIEVHWYTRK ANLHHIWDSN  420
IIETAEADLY NSALEGMVDA LKKNITTEWA DQVKRWETCT KKTACPDIYA SEGIQAACDW  480
AYKGVTEGDT LEDEYFYSRL PIVYQRLAQG GVRLAATLNR IFGGSSGEKI KGLNEYINLY  540
NQKTKQKLPK FKPLYKQVLS DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE  600
KLFKNFDEYS SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE  660
DDRRKSFKKI GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL  720
EKSLKKNDAV VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH  780
IYDAIRNYVT QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY  840
AKCLQKIDKD DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK  900
KGDMFNLNDC HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF  960
ESASKKEVDK LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA  1020
ELFMRRASLK KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI  1080
NKCPKNIFKI NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE QYSLNEIINN  1140
FNGIRIKTDY HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI  1200
ALEDLNSGFK NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES  1260
FKSMSTQNGF IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL  1320
FEFALDYKNF SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY  1380
GINYQQGDIR ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY  1440
DSRNYEAQEN AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA  1500
QTSVKH                                                             1506

SEQ ID NO: 252          moltype = AA  length = 1320
FEATURE                 Location/Qualifiers
source                  1..1320
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 252
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGIDGVKLS STEKKKMDEI APPKGSVFTC PICEKRSIVG VTANLVHDHN  300
HDTGWGREWI CDSCNTGLGR FKDNPKFLEK VIEYLKKGSS GEKIKGLNEY INLYNQKTKQ  360
KLPKFKPLYK QVLSDRESLS FYGEGYTSDE EVLEVFRNTL NKNSEIFSSI KKLEKLFKNF  420
DEYSSAGIFV KNGPAISTIS KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT EKYEDDRRKS  480
FKKIGSFSLE QLQEYADADL SVVEKLKEII IQKVDEIYKV YGSSEKLFDA DFVLEKSLKK  540
NDAVVAIMKD LLDSVKSFEN YIKAFFGEGK ETNRDESFYG DFVLAYDILL KVDHIYDAIR  600
NYVTQKPYSK DKFKLYFQNP QFMGGWDKDK ETDYRATILR YGSKYYLAIM DKKYAKCLQK  660
IDKDDVNGNY EKINYKLLPG PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFN  720
LNDCHKLIDF FKDSISRYPK WSNAYDFNFS ETEKYKDIAG FYREVEEQGY KVSFESASKK  780
EVDKLVEEGK LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF DENNHGQIRL SGGAELFMRR  840
ASLKKEELVV HPANSPIANK NPDNPKKTTT LSYDVYKDKR FSEDQYELHI PIAINKCPKN  900
IFKINTEVRV LLKHDDNPYV IGIARGERNL LYIVVVDGKG NIVEQYSLNE IINNFNGIRI  960
KTDYHSLLDK KEKERFEARQ NWTSIENIKE LKAGYISQVV HKICELVEKY DAVIALEDLN  1020
SGFKNSRVKV EKQVYQKFEK MLIDKLNYMV DKKSNPCATG GALKGYQITN KFESFKSMST  1080
QNGFIFYIPA WLTSKIDPST GFVNLLKTKY TSIADSKKFI SSFDRIMYVP EEDLFEFALD  1140
YKNFSRTDAD YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL FNKYGINYQQ  1200
GDIRALLCEQ SDKAFYSSFM ALMSLMLQMR NSITGRTDVD FLISPVKNSD GIFYDSRNYE  1260
AQENAILPKN ADANGAYNIA RKVLWAIGQF KKAEDEKLDK VKIAISNKEW LEYAQTSVKH  1320

SEQ ID NO: 253          moltype = AA  length = 1333
FEATURE                 Location/Qualifiers
source                  1..1333
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 253
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGMNYFIVE VSEQEVKREK EKARELRRSQ WWKNRIARGI CHYCGEIFPP  300
EELTMDHLVP VVRGGKSTRG NVVPACKECN NRKKYLLPVE WEEYLDSLES GSSGEKIKGL  360
NEYINLYNQK TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF  420
SSIKKLEKLF KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA  480
VVTEKYEDDR RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL  540
FDADFVLEKS LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD  600
ILLKVDHIYD AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL  660
AIMDKKYAKC LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI  720
YKNGTFKKGD MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE  780
QGYKVSFESA SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ  840
```

-continued

```
IRLSGGAELF MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE  900
LHIPIAINKC PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS  960
LNEIINNFNG IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV  1020
EKYDAVIALE DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ  1080
ITNKFESFKS MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM  1140
YVPEEDLFEF ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY  1200
KELFNKYGIN YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK  1260
NSDGIFYDSR NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN  1320
KEWLEYAQTS VKH                                                     1333

SEQ ID NO: 254          moltype = AA  length = 1496
FEATURE                 Location/Qualifiers
source                  1..1496
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 254
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGGKAQLDI KNFPELYRTT ERVYKKSGQS TKPVTVSNIH YSVLDGYGRS  300
GEAYGIITKD MIDMSAGYRE KWESKPEPSG WYSYFFKNTN QRATESDYKH SPKNVSKISN  360
NIKASILLSN GNVRNGYLFD RSHLIADSLG GRPFRNNLIT GTRTQNVGNN DRKGGMQYIE  420
NKVLDHIKRN PKVHVYYKAT PVYQGSELLP RAVLVSALSS DGFIDETVRV FNNVAGFNID  480
YQNGGLLSST ADVDINNVEE NEIETTDDEI EEGGSSGEKI KGLNEYINLY NQKTKQKLPK  540
FKPLYKQVLS DRESLSFYGE GYTSDEEVLE VFRNTLNKNS EIFSSIKKLE KLFKNFDEYS  600
SAGIFVKNGP AISTISKDIF GEWNVIRDKW NAEYDDIHLK KKAVVTEKYE DDRRKSFKKI  660
GSFSLEQLQE YADADLSVVE KLKEIIIQKV DEIYKVYGSS EKLFDADFVL EKSLKKNDAV  720
VAIMKDLLDS VKSFENYIKA FFGEGKETNR DESFYGDFVL AYDILLKVDH IYDAIRNYVT  780
QKPYSKDKFK LYFQNPQFMG GWDKDKETDY RATILRYGSK YYLAIMDKKY AKCLQKIDKD  840
DVNGNYEKIN YKLLPGPNKM LPKVFFSKKW MAYYNPSEDI QKIYKNGTFK KGDMFNLNDC  900
HKLIDFFKDS ISRYPKWSNA YDFNFSETEK YKDIAGFYRE VEEQGYKVSF ESASKKEVDK  960
LVEEGKLYMF QIYNKDFSDK SHGTPNLHTM YFKLLFDENN HGQIRLSGGA ELFMRRASLK  1020
KEELVVHPAN SPIANKNPDN PKKTTTLSYD VYKDKRFSED QYELHIPIAI NKCPKNIFKI  1080
NTEVRVLLKH DDNPYVIGIA RGERNLLYIV VVDGKGNIVE QYSLNEIINN FNGIRIKTDY  1140
HSLLDKKEKE RFEARQNWTS IENIKELKAG YISQVVHKIC ELVEKYDAVI ALEDLNSGFK  1200
NSRVKVEKQV YQKFEKMLID KLNYMVDKKS NPCATGGALK GYQITNKFES FKSMSTQNGF  1260
IFYIPAWLTS KIDPSTGFVN LLKTKYTSIA DSKKFISSFD RIMYVPEEDL FEFALDYKNF  1320
SRTDADYIKK WKLYSYGNRI RIFRNPKKNN VFDWEEVCLT SAYKELFNKY GINYQQGDIR  1380
ALLCEQSDKA FYSSFMALMS LMLQMRNSIT GRTDVDFLIS PVKNSDGIFY DSRNYEAQEN  1440
AILPKNADAN GAYNIARKVL WAIGQFKKAE DEKLDKVKIA ISNKEWLEYA QTSVKH      1496

SEQ ID NO: 255          moltype = AA  length = 1380
FEATURE                 Location/Qualifiers
source                  1..1380
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 255
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGENQTTGK GKNNSRPRYK SLEKAIKEFG SQILKEHPTD NQELRNNRLY  300
LYYLQNGKDM YTGQDLDIHN LSNYDIDHIV PQSFITDNSI DNLVLTSSAG NREKGDDVPP  360
LEIVRKRKVF WEKLYQGNLM SKRKFDYLTK AERGGLTGSS GEKIKGLNEY INLYNQKTKQ  420
KLPKFKPLYK QVLSDRESLS FYGEGYTSDE EVLEVFRNTL NKNSEIFSSI KKLEKLFKNF  480
DEYSSAGIFV KNGPAISTIS KDIFGEWNVI RDKWNAEYDD IHLKKKAVVT EKYEDDRRKS  540
FKKIGSFSLE QLQEYADADL SVVEKLKEII IQKVDEIYKV YGSSEKLFDA DFVLEKSLKK  600
NDAVVAIMKD LLDSVKSFEN YIKAFFGEGK ETNRDESFYG DFVLAYDILL KVDHIYDAIR  660
NYVTQKPYSK DKFKLYFQNP QFMGGWDKDK ETDYRATILR YGSKYYLAIM DKKYAKCLQK  720
IDKDDVNGNY EKINYKLLPG PNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFN  780
LNDCHKLIDF FKDSISRYPK WSNAYDFNFS ETEKYKDIAG FYREVEEQGY KVSFESASKK  840
EVDKLVEEGK LYMFQIYNKD FSDKSHGTPN LHTMYFKLLF DENNHGQIRL SGGAELFMRR  900
ASLKKEELVV HPANSPIANK NPDNPKKTTT LSYDVYKDKR FSEDQYELHI PIAINKCPKN  960
IFKINTEVRV LLKHDDNPYV IGIARGERNL LYIVVVDGKG NIVEQYSLNE IINNFNGIRI  1020
KTDYHSLLDK KEKERFEARQ NWTSIENIKE LKAGYISQVV HKICELVEKY DAVIALEDLN  1080
SGFKNSRVKV EKQVYQKFEK MLIDKLNYMV DKKSNPCATG GALKGYQITN KFESFKSMST  1140
QNGFIFYIPA WLTSKIDPST GFVNLLKTKY TSIADSKKFI SSFDRIMYVP EEDLFEFALD  1200
YKNFSRTDAD YIKKWKLYSY GNRIRIFRNP KKNNVFDWEE VCLTSAYKEL FNKYGINYQQ  1260
GDIRALLCEQ SDKAFYSSFM ALMSLMLQMR NSITGRTDVD FLISPVKNSD GIFYDSRNYE  1320
AQENAILPKN ADANGAYNIA RKVLWAIGQF KKAEDEKLDK VKIAISNKEW LEYAQTSVKH  1380

SEQ ID NO: 256          moltype = AA  length = 1323
FEATURE                 Location/Qualifiers
source                  1..1323
                        mol_type = protein
```

-continued

```
                      organism = synthetic construct
SEQUENCE: 256
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGMELKNSI SDYTEAEFVQ LLKEIEKENV AATDDVLDVL LEHFVKITEH  300
PDGTDLIYYP SDNRDDSPEG IVKEIKEWRA ANGKPGFKQG GSSGEKIKGL NEYINLYNQK  360
TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF  420
KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR  480
RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS  540
LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD  600
AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC  660
LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD  720
MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA  780
SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF  840
MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC  900
PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG  960
IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE  1020
DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS  1080
MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF  1140
ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN  1200
YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR  1260
NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS  1320
VKH                                                                1323

SEQ ID NO: 257        moltype = AA  length = 1323
FEATURE               Location/Qualifiers
source                1..1323
                      mol_type = protein
                      organism = synthetic construct
SEQUENCE: 257
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGMKSKISE YTEKEFLEFV KDIYTNNKKK FPTEESHIQA VLEFKKLTEH  300
PSGSDLLYYP NENREDSPAG VVKEVKEWRA SKGLPGFKAG GSSGEKIKGL NEYINLYNQK  360
TKQKLPKFKP LYKQVLSDRE SLSFYGEGYT SDEEVLEVFR NTLNKNSEIF SSIKKLEKLF  420
KNFDEYSSAG IFVKNGPAIS TISKDIFGEW NVIRDKWNAE YDDIHLKKKA VVTEKYEDDR  480
RKSFKKIGSF SLEQLQEYAD ADLSVVEKLK EIIIQKVDEI YKVYGSSEKL FDADFVLEKS  540
LKKNDAVVAI MKDLLDSVKS FENYIKAFFG EGKETNRDES FYGDFVLAYD ILLKVDHIYD  600
AIRNYVTQKP YSKDKFKLYF QNPQFMGGWD KDKETDYRAT ILRYGSKYYL AIMDKKYAKC  660
LQKIDKDDVN GNYEKINYKL LPGPNKMLPK VFFSKKWMAY YNPSEDIQKI YKNGTFKKGD  720
MFNLNDCHKL IDFFKDSISR YPKWSNAYDF NFSETEKYKD IAGFYREVEE QGYKVSFESA  780
SKKEVDKLVE EGKLYMFQIY NKDFSDKSHG TPNLHTMYFK LLFDENNHGQ IRLSGGAELF  840
MRRASLKKEE LVVHPANSPI ANKNPDNPKK TTTLSYDVYK DKRFSEDQYE LHIPIAINKC  900
PKNIFKINTE VRVLLKHDDN PYVIGIARGE RNLLYIVVVD GKGNIVEQYS LNEIINNFNG  960
IRIKTDYHSL LDKKEKERFE ARQNWTSIEN IKELKAGYIS QVVHKICELV EKYDAVIALE  1020
DLNSGFKNSR VKVEKQVYQK FEKMLIDKLN YMVDKKSNPC ATGGALKGYQ ITNKFESFKS  1080
MSTQNGFIFY IPAWLTSKID PSTGFVNLLK TKYTSIADSK KFISSFDRIM YVPEEDLFEF  1140
ALDYKNFSRT DADYIKKWKL YSYGNRIRIF RNPKKNNVFD WEEVCLTSAY KELFNKYGIN  1200
YQQGDIRALL CEQSDKAFYS SFMALMSLML QMRNSITGRT DVDFLISPVK NSDGIFYDSR  1260
NYEAQENAIL PKNADANGAY NIARKVLWAI GQFKKAEDEK LDKVKIAISN KEWLEYAQTS  1320
VKH                                                                1323

SEQ ID NO: 258        moltype = AA  length = 1511
FEATURE               Location/Qualifiers
source                1..1511
                      mol_type = protein
                      organism = synthetic construct
SEQUENCE: 258
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGWRRRRED ARAAPGLLGR LPVLPVAAAA ELPPVPGGPR GPGELAKYGL  300
PGLAQLKSRE SYVLCYDPRT RGALWVVEQL RPERLRGDGD RRECDFREDD SVHAYHRATN  360
ADYRGSGFDR GHLAAAANHR WSQKAMDDTF YLSNVAPQVP HLNQNAWNNL EKYSRSLTRS  420
YQNVYVCTGP LFLPRTEADG KSYVKYQVIG KNHVAVPTHF FKVLILEAAG GQIELRTYVM  480
PNAPVDEAIP LERFLVPIES IERASGLLFV PNILARAGSL KAITAGSKGS SGEKIKGLNE  540
YINLYNQKTK QKLPKFKPLY KQVLSDRESL SFYGEGYTSD EEVLEVFRNT LNKNSEIFSS  600
IKKLEKLFKN FDEYSSAGIF VKNGPAISTI SKDIFGEWNV IRDKWNAEYD DIHLKKKAVV  660
TEKYEDDRRK SFKKIGSFSL EQLQEYADAD LSVVEKLKEI IIQKVDEIYK VYGSSEKLFD  720
ADFVLEKSLK KNDAVVAIMK DLLDSVKSFE NYIKAFFGEG KETNRDESFY GDFVLAYDIL  780
LKVDHIYDAI RNYVTQKPYS KDKFKLYFQN PQFMGGWDKD KETDYRATIL RYGSKYYLAI  840
```

-continued

```
MDKKYAKCLQ KIDKDDVNGN YEKINYKLLP GPNKMLPKVF FSKKWMAYYN PSEDIQKIYK  900
NGTFKKGDMF NLNDCHKLID FFKDSISRYP KWSNAYDFNF SETEKYKDIA GFYREVEEQG  960
YKVSFESASK KEVDKLVEEG KLYMFQIYNK DFSDKSHGTP NLHTMYFKLL FDENNHGQIR  1020
LSGGAELFMR RASLKKEELV VHPANSPIAN KNPDNPKKTT TLSYDVYKDK RFSEDQYELH  1080
IPIAINKCPK NIFKINTEVR VLLKHDDNPY VIGIARGERN LLYIVVVDGK GNIVEQYSLN  1140
EIINNFNGIR IKTDYHSLLD KKEKERFEAR QNWTSIENIK ELKAGYISQV VHKICELVEK  1200
YDAVIALEDL NSGFKNSRVK VEKQVYQKFE KMLIDKLNYM VDKKSNPCAT GGALKGYQIT  1260
NKFESFKSMS TQNGFIFYIP AWLTSKIDPS TGFVNLLKTK YTSIADSKKF ISSFDRIMYV  1320
PEEDLFEFAL DYKNFSRTDA DYIKKWKLYS YGNRIRIFRN PKKNNVFDWE EVCLTSAYKE  1380
LFNKYGINYQ QGDIRALLCE QSDKAFYSSF MALMSLMLQM RNSITGRTDV DFLISPVKNS  1440
DGIFYDSRNY EAQENAILPK NADANGAYNI ARKVLWAIGQ FKKAEDEKLD KVKIAISNKE  1500
WLEYAQTSVK H                                                       1511

SEQ ID NO: 259          moltype = AA  length = 1372
FEATURE                 Location/Qualifiers
source                  1..1372
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 259
MSKLEKFTNC YSLSKTLRFK AIPVGKTQEN IDNKRLLVED EKRAEDYKGV KKLLDRYYLS  60
FINDVLHSIK LKNLNNYISL FRKKTRTEKE NKELENLEIN LRKEIAKAFK GNEGYKSLFK  120
KDIIETILPE FLDDKDEIAL VNSFNGFTTA FTGFFDNREN MFSEEAKSTS IAFRCINENL  180
TRYISNMDIF EKVDAIFDKH EVQEIKEKIL NSDYDVEDFF EGEFFNFVLT QEGIDVYNAI  240
IGGFVTESGG SSGDEPGVAT GNGQPVTGNW LAGASQGDGV PIPSQIADQL RGKEFKSWRD  300
FREQFWMAVS KDPSALENLS PSNRYFVSQG LAPYAVPEEH LGSKEKFEIH HVVPLESGGA  360
LYNIDNLVIV TPKRHSEIHK ELKLKRKEKG SSGEKIKGLN EYINLYNQKT KQKLPKFKPL  420
YKQVLSDRES LSFYGEGYTS DEEVLEVFRN TLNKNSEIFS SIKKLEKLFK NFDEYSSAGI  480
FVKNGPAIST ISKDIFGEWN VIRDKWNAEY DDIHLKKKAV VTEKYEDDRR KSFKKIGSFS  540
LEQLQEYADA DLSVVEKLKE IIIQKVDEIY KVYGSSEKLF DADFVLEKSL KKNDAVVAIM  600
KDLLDSVKSF ENYIKAFFGE GKETNRDESF YGDFVLAYDI LLKVDHIYDA IRNYVTQKPY  660
SKDKFKLYFQ NPQFMGGWDK DKETDYRATI LRYGSKYYLA IMDKKYAKCL QKIDKDDVNG  720
NYEKINYKLL PGPNKMLPKV FFSKKWMAYY NPSEDIQKIY KNGTFKKGDM FNLNDCHKLI  780
DFFKDSISRY PKWSNAYDFN FSETEKYKDI AGFYREVEEQ GYKVSFESAS KKEVDKLVEE  840
GKLYMFQIYN KDFSDKSHGT PNLHTMYFKL LFDENNHGQI RLSGGAELFM RRASLKKEEL  900
VVHPANSPIA NKNPDNPKKT TTLSYDVYKD KRFSEDQYEL HIPIAINKCP KNIFKINTEV  960
RVLLKHDDNP YVIGIARGER NLLYIVVVDG KGNIVEQYSL NEIINNFNGI RIKTDYHSLL  1020
DKKEKERFEA RQNWTSIENI KELKAGYISQ VVHKICELVE KYDAVIALED LNSGFKNSRV  1080
KVEKQVYQKF EKMLIDKLNY MVDKKSNPCA TGGALKGYQI TNKFESFKSM STQNGFIFYI  1140
PAWLTSKIDP STGFVNLLKT KYTSIADSKK FISSFDRIMY VPEEDLFEFA LDYKNFSRTD  1200
ADYIKKWKLY SYGNRIRIFR NPKKNNVFDW EEVCLTSAYK ELFNKYGINY QQGDIRALLC  1260
EQSDKAFYSS FMALMSLMLQ MRNSITGRTD VDFLISPVKN SDGIFYDSRN YEAQENAILP  1320
KNADANGAYN IARKVLWAIG QFKKAEDEKL DKVKIAISNK EWLEYAQTSV KH          1372
```

1.-**10**. (canceled)

11. The composition of claim **23**, wherein the first fusion protein and/or the second fusion protein further comprises a reverse transcriptase.

12. The composition of claim **23**, wherein the first Cas12a polypeptide and/or the second Cas12a polypeptide comprises about 175, 310, 406, 441, or 550 consecutive amino acids of a Cas12a protein.

13. The composition of claim **23**, wherein the first Cas12a polypeptide and/or the second Cas12a polypeptide has at least 70% sequence identity to one or more of SEQ ID NOs:150-159 and 175-184.

14. The composition of claim **23**, wherein the first fusion protein and/or the second fusion protein has at least 70% sequence identity to one or more of SEQ ID NOs:100-109 or 188.

15.-**22**. (canceled)

23. A composition comprising:

a first fusion protein that comprises a first Cas12a polypeptide fused to a first intein polypeptide; and

a second fusion protein that comprises a second Cas12a polypeptide fused to a second intein polypeptide.

24. The composition of claim **23**, wherein the first fusion protein and the second fusion protein are different.

25.-**28**. (canceled)

29. A method of modifying a target nucleic acid, the method comprising:

introducing a first nucleic acid molecule and a second nucleic acid molecule into a cell comprising the target nucleic acid, wherein the first nucleic acid molecule encodes a first fusion protein, the first fusion protein comprising a first Cas12a polypeptide fused to a first intein polypeptide and the second nucleic acid molecule encodes a second fusion protein, the second fusion protein comprising a second Cas12a polypeptide fused to a second intein polypeptide;

contacting the target nucleic acid in the cell with a guide nucleic acid and a protein comprising at least a portion of the first Cas12a polypeptide and at least a portion of the second Cas12a polypeptide,

optionally wherein the protein and the guide nucleic acid form a complex or are comprised in a complex, thereby modifying the target nucleic acid.

30.-**32**. (canceled)

33. The method of claim **29**, wherein the first intein polypeptide and the second polypeptide associate to form an intein that is a *Nostoc punctiforme* (Npu) intein and/or a portion of a mutant Npu intein.

34. The method of claim **29**, wherein the first intein polypeptide and/or the second intein polypeptide has at least 70% sequence identity to one or more of SEQ ID NOs:110-112.

35. (canceled)

36. The method of claim **29**, wherein the protein has at least 70% sequence identity to one or more of SEQ ID NOs:38-60, 113-149, 192-195, and 196-259.

37. The method of claim **29**, wherein the first fusion protein has at least 70% sequence identity to one of SEQ ID NOs:100, 102, 104, 106, and 108 and the second fusion protein has at least 70% sequence identity to one of SEQ ID NOs:101, 103, 105, 107, and 109, respectively, wherein the first fusion protein and the second fusion protein are different.

38.-**41**. (canceled)

42. An engineered protein comprising an amino acid sequence that has at least 70% sequence identity to one or more of SEQ ID NOs:100-109 and 187-188.

43. A nucleic acid molecule encoding the engineered protein of claim **42**.

44.-**69**. (canceled)

70. A method of modifying a target nucleic acid, the method comprising:

introducing a first nucleic acid molecule and a second nucleic acid molecule into a cell comprising the target nucleic acid, wherein the first nucleic acid molecule encodes a first fusion protein, the first fusion protein comprising a polypeptide of interest fused to a first intein polypeptide and the second nucleic acid molecule encodes a second fusion protein, the second fusion protein comprising a Cas12a polypeptide fused to a second intein polypeptide;

contacting the target nucleic acid in the cell with a guide nucleic acid (e.g., a guide RNA) and a protein comprising at least a portion of the polypeptide of interest and at least a portion of the Cas12a polypeptide,

optionally wherein the protein and the guide nucleic acid form a complex or are comprised in a complex, thereby modifying the target nucleic acid.

71.-**73**. (canceled)

74. The method of claim **70**, wherein the first intein polypeptide and the second polypeptide associate to form an intein that is a *Nostoc punctiforme* (Npu) intein and/or a portion of a mutant Npu intein.

75. The method of claim **70**, wherein the first intein polypeptide and/or the second polypeptide has at least 70% sequence identity to one or more of SEQ ID NOs:110-112.

76. The method of claim **70**, further comprising cleaving the polypeptide of interest from the first fusion protein and cleaving the Cas12a polypeptide from the second fusion protein.

77. The method of claim **76**, wherein the fusion protein has at least 70% sequence identity to one or more of SEQ ID NOs:38-60, 113-149, 192-195, and 196-259.

78. The method of claim **70**, wherein the first fusion protein has at least 70% sequence identity to SEQ ID NO:187 and the second fusion protein has at least 70% sequence identity to SEQ ID NO:188.

79.-**81**. (canceled)

82. The method of claim **70**, wherein the method has increased efficiency in modifying the target nucleic acid compared to the efficiency of a control method.

* * * * *