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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
06 May 2021 (06.05.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/087182 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/55* (2006.01)
C12N 15/11 (2006.01) *C12N 15/63* (2006.01)
C12N 15/113 (2010.01) *C12N 15/90* (2006.01)

(21) International Application Number:
PCT/US2020/058082

(22) International Filing Date:
30 October 2020 (30.10.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/927,914 30 October 2019 (30.10.2019) US

(71) Applicant: **PAIRWISE PLANTS SERVICES, INC.** [US/US]; 807 East Main Street, Suite 4-100, Durham, North Carolina 27701 (US).

(72) Inventor: **KIM, Yongjoo**; 605 W Chapel Hill Street, Apt. 704, Durham, North Carolina 27701 (US).

(74) Agent: **BOBAY, Erin Regel**; MYERS BIGEL, P.A., PO Box 37428, Raleigh, North Carolina 27627 (US).

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

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

(54) Title: TYPE V CRISPR-CAS BASE EDITORS AND METHODS OF USE THEREOF

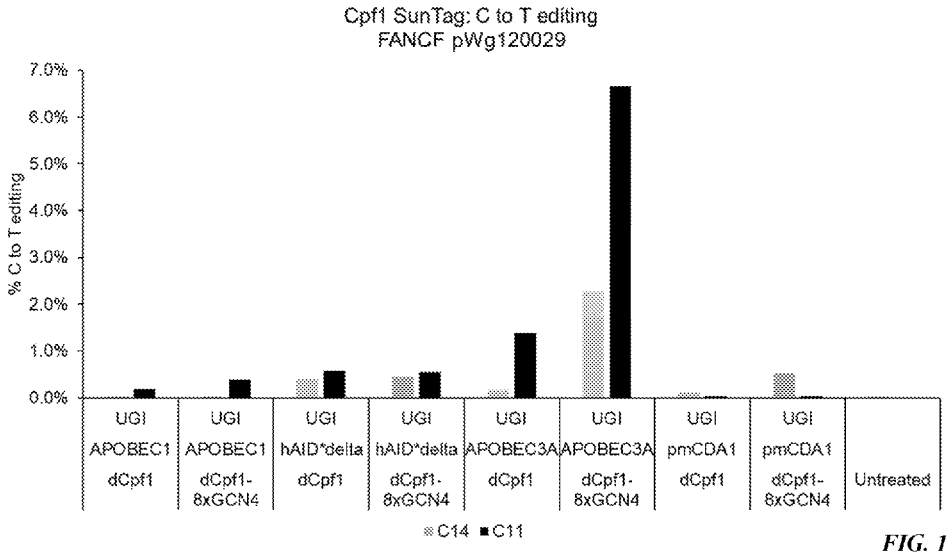


FIG. 1

(57) Abstract: This invention relates to Type V CRISPR-Cas effector proteins, deaminases, and fusion and recruiting nucleic acid constructs thereof. The invention further relates methods of targeted nucleic acid modification utilizing the same. One aspect of the invention provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas effector protein; (b) a deaminase, optionally wherein the target nucleic acid is contacted with two or more deaminase; and(c) a guide nucleic acid, wherein the deaminase is recruited to the Type V CRISPR-Cas effector protein (e.g., recruited via a protein to protein interaction, RNA to protein interaction, and/or chemical interaction), thereby modifying the target nucleic acid, optionally wherein the Type V CRISPR-Cas effector protein, the deaminase and guide nucleic acid are co-expressed.

WO 2021/087182 A1

[Continued on next page]

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

Published:

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

TYPE V CRISPR-CAS BASE EDITORS AND METHODS OF USE THEREOF

STATEMENT REGARDING ELECTRONIC FILING OF A SEQUENCE LISTING

A Sequence Listing in ASCII text format, submitted under 37 C.F.R. § 1.821, entitled 1499-10WO_ST25.txt, 351,999 bytes in size, generated on October 30, 2020 and filed via EFS-Web, is provided in lieu of a paper copy. This Sequence Listing is hereby incorporated herein by reference into the specification for its disclosures.

FIELD OF THE INVENTION

This invention relates to Type V CRISPR-Cas effector proteins, deaminases, and fusion and recruiting nucleic acid constructs thereof. The invention further relates methods of targeted nucleic acid modification utilizing the same.

BACKGROUND OF THE INVENTION

Gene editing is the process of utilizing a site-directed nuclease to introduce variation at targeted genomic locations. C to T base editing may be achieved with base editors that use Cas9 as the targeting module. For example, Komor et al. (*Sci Advances* 3(8):eaao4774) (2017)) and Koblan et al. (*Nat Biotechnol* 36(9):843-846 (2018) disclose base editors that use Cas9 as a fusion protein including APOBEC1 deaminase, Cas9 nickase (D10A), and 2 copies of Uracil glycosylase inhibitor (UGI). Li et al. (*Nat Biotechnol*. 36(4):324-327 (2018)) replaces Cas9 with deactivated Cpf1. Although active in human cells, this Cpf1 construct is less efficient than its Cas9 counterparts. To make base editing more useful across a greater number of organisms, including plants, new base editing tools are needed.

SUMMARY OF THE INVENTION

One aspect of the invention provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas effector protein; (b) a deaminase, optionally wherein the target nucleic acid is contacted with two or more deaminase; and(c) a guide nucleic acid, wherein the deaminase is recruited to the Type V CRISPR-Cas effector protein (e.g., recruited via a protein to protein interaction, RNA to protein interaction, and/or chemical interaction), thereby modifying the target nucleic acid,

optionally wherein the Type V CRISPR-Cas effector protein, the deaminase and guide nucleic acid are co-expressed.

A second aspect of the invention provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag (e.g., an epitope or a multimerized epitope); (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag, optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and (c) a guide nucleic acid, optionally wherein the Type V CRISPR-Cas fusion protein, the deaminase fusion protein and guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

A third aspect of the invention provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to an affinity polypeptide that binds to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to the peptide tag (e.g., an epitope or a multimerized epitope), optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and (c) a guide nucleic acid, optionally wherein the Type V CRISPR-Cas fusion protein, the deaminase fusion protein and guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

A fourth aspect provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas effector protein; (b) a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif, and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and optionally wherein the Type V CRISPR-Cas effector protein, the deaminase fusion protein and the recruiting guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

A fifth aspect of the invention provides a nucleic acid construct comprising: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and (c) a guide nucleic acid.

A sixth aspect of the invention provides a nucleic acid construct comprising: (a) a Type V CRISPR-Cas effector protein; (b) a recruiting guide nucleic acid comprising a guide

nucleic acid linked to an RNA recruiting motif; and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif.

A seventh aspect of the invention provides a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system comprising: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and (c) a guide nucleic acid comprising a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Type V CRISPR-Cas effector protein of the Type V CRISPR-Cas fusion protein and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the Type V CRISPR-Cas fusion protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Type V CRISPR-Cas fusion protein and target nucleic acid by the binding of the affinity polypeptide to the peptide tag that is fused to the Type V CRISPR-Cas fusion protein, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid.

A eighth aspect of the invention provides a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system comprising: (a) a Type V CRISPR-Cas effector protein; (b) a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif, and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, wherein the recruiting guide nucleic acid comprises a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Type V CRISPR-Cas effector protein and the recruiting guide nucleic acid is capable of hybridizing to a target nucleic acid, thereby guiding the Type V CRISPR-Cas effector protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Type V CRISPR-Cas effector protein and target nucleic acid by the binding of the affinity polypeptide to the RNA recruiting motif that is fused to the recruiting guide nucleic acid, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid.

The invention further provides expression cassettes and/or vectors comprising the nucleic acid constructs of the invention, and cells comprising the polypeptides, fusion proteins and/or nucleic acid constructs of the invention. Additionally, the invention provides kits comprising the nucleic acid constructs of the invention and expression cassettes, vectors and/or cells comprising the same.

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

Fig. 1 is a graph showing C to T editing efficiencies for editing systems using pWg120029 as the guide nucleic acid according to some embodiments of the present invention.

Fig. 2 is a graph showing C to T editing efficiencies for editing systems using pWg120360 as the guide nucleic acid according to some embodiments of the present invention.

Fig. 3 is a graph showing C to T editing efficiencies for editing systems using pWg120300 as the guide nucleic acid according to some embodiments of the present invention.

Fig. 4 is a graph showing C to T editing efficiencies for editing systems using pWg120301 as the guide nucleic acid according to some embodiments of the present invention.

Fig. 5 is a graph showing A to G editing efficiencies for editing systems using different the guide nucleic acids according to some embodiments of the present invention.

DETAILED DESCRIPTION

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.

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.

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.

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.

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.

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").

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 measureable value may include any other range and/or individual value therein.

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."

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.

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.

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."

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 as compared to a control.

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% as compared to a control. In particular 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.

A "heterologous" 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.

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 "wild type mRNA" is an mRNA that is naturally occurring in or endogenous to the reference organism. A "homologous" nucleic acid sequence is a nucleotide sequence naturally associated with a host cell into which it is introduced.

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.

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.

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). A gene may be "isolated" by which is meant a nucleic acid that is substantially or essentially free from components normally found in association with the nucleic acid in its natural state. Such components include other cellular material, culture medium from recombinant production, and/or various chemicals used in chemically synthesizing the nucleic acid.

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.

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.

"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," 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%, and the like, complementarity).

A “portion” or “fragment” of a nucleotide sequence or polypeptide 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 contiguous residues 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. 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).

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 polypeptide sequences of this invention. "Orthologous," 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 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.

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).

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.

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.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences 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.

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). 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 polynucleotide sequences may be to a full-length polynucleotide sequence or a portion thereof, or to a longer polynucleotide sequence. For purposes of this invention "percent identity" may also be determined using BLASTX version 2.0 for translated nucleotide sequences and BLASTN version 2.0 for polynucleotide sequences.

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.

"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.

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 0.2x 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 1x 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 4-6x 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 2x (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.

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 invention (comprising/encoding a Type V CRISPR-Cas effector protein, a deaminase fusion protein, and a recruiting guide nucleic acid or a Type V CRISPR-Cas fusion protein, a deaminase fusion protein, and a guide nucleic acid) may be codon optimized for expression in an organism (e.g., an animal, 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.

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 a an organism or cell thereof (e.g., a plant and/or a cell of a plant). 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).

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.

As used herein, the term "linked," or "fused" in reference to polypeptides, refers to the attachment of one polypeptide to another. A polypeptide may be linked or fused to another polypeptide (at the N-terminus or the C-terminus) directly (e.g., via a peptide bond) or through a linker (e.g., a peptide linker).

The term "linker" in reference to polypeptides is art-recognized and refers to a chemical group, or a molecule linking two molecules or moieties, e.g., two polypeptides (e.g., domains) of a fusion protein, such as, for example, a Type V CRISPR-Cas effector protein and a peptide tag and/or a polypeptide of interest. 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.

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 be a GS linker.

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 covalent or non-covenant linkage or binding, including e.g., Watson-Crick base-pairing, 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.

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:1** or **SEQ ID NO:2**).

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.

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.

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)).

Examples of constitutive promoters useful for plants include, but are not limited to, cestrum virus promoter (cmp) (U.S. Patent No. 7,166,770), the rice actin 1 promoter (Wang *et al.* (1992) *Mol. Cell. Biol.* 12:3399-3406; as well as US Patent 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.

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 β-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 US Patent 6,040,504; the rice sucrose synthase promoter disclosed in US Patent 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. Patent 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. Patent No. 10,421,972), PLA_2-δ promoter from arabidopsis (U.S. Patent No. 7,141,424), and/or the ZmC5 promoter from maize (International PCT Publication No. WO1999/042587.

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. Patent No. 5459252), 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-l,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).

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. Patent 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).

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. Patent 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).

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

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.

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.

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 Type V CRISPR-Cas effector protein , a polynucleotide encoding a Type V CRISPR-Cas fusion protein, a polynucleotide encoding a deaminase (e.g., a cytosine deaminase and/or an adenine deaminase), a polynucleotide encoding a deaminase fusion protein, a polynucleotide encoding a peptide tag, a polynucleotide encoding an affinity polypeptide, a recruiting guide

nucleic acid and/or 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 different promoters (e.g., three polynucleotides may be driven by one, two or three promoters in any combination). Thus, for example, a polynucleotide encoding a Type V CRISPR-Cas fusion protein, a polynucleotide encoding a deaminase fusion protein, and a guide nucleic acid comprised in an expression cassette may each be operably associated with a single promoter or they may be operably associated with separate promoters (e.g., two or three promoters) in any combination. As another example, a polynucleotide encoding a Type V CRISPR-Cas effector protein, a polynucleotide encoding a deaminase fusion protein, and a recruiting guide nucleic acid comprised in an expression cassette may each be operably associated with a single promoter or they may be operably associated with separate promoters (e.g., two or three promoters) in any combination.

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).

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.

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).

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.

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 comprises a nucleic acid construct comprising the nucleotide sequence(s) to be transferred, delivered or introduced. 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, 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.

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, Type V CRISPR-Cas fusion protein, a deaminase fusion protein and a guide nucleic acid, under conditions whereby the Type V CRISPR-Cas fusion protein is expressed, and the Type V CRISPR-Cas fusion protein forms a complex with the guide nucleic acid, the complex hybridizes to the target nucleic acid, and the deaminase fusion protein is recruited to the Type V CRISPR-Cas effector protein (and thus, to the target nucleic acid), thereby modifying the target nucleic acid. In some embodiments, a Type V CRISPR-Cas protein (optionally a Type V CRISPR-Cas fusion protein), a guide nucleic acid, and a deaminase (optionally a deaminase fusion protein) contact a target nucleic acid to thereby modify the nucleic acid. In some embodiments, the Type V CRISPR-Cas protein, guide nucleic acid, and/or deaminase may be in the form of a complex (e.g., a ribonucleoprotein such as an assembled ribonucleoprotein complex) and the complex contacts the target nucleic acid. In some embodiments, the complex or a component thereof (e.g., the guide nucleic acid) hybridizes to the target nucleic acid and thereby the target nucleic acid is modified (e.g., via action of the Type V CRISPR-Cas protein and/or deaminase). In some embodiments, a deaminase or deaminase fusion protein and the Type V CRISPR-Cas effector protein localize at the target nucleic acid, optionally through covalent and/or non-covalent interactions.

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, nicking, and/or transcriptional control of a target nucleic acid.

"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, RNA-protein interactions, and/or chemical interactions. Protein-protein interactions can include, but are not limited to, peptide tags (e.g., 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) (Voß 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)).

"Introducing," "introduce," "introduced" (and grammatical variations thereof) in the context of a polynucleotide of interest means presenting a nucleotide sequence of interest (e.g., polynucleotide, a nucleic acid construct, and/or a guide nucleic acid) 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 gains access to the interior of a cell. Thus, for example, a nucleic acid construct of the invention encoding a Type V CRISPR-Cas fusion protein and a deaminase fusion protein as described herein and guide nucleic acid may be introduced into a cell of an organism, thereby transforming the cell with the Type V CRISPR-Cas effector protein fusion protein, the deaminase fusion protein and the guide nucleic acid.

The term "transformation" as used herein refers to the introduction of a heterologous nucleic acid 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 of the invention.

"Transient transformation" in the context of a polynucleotide means that a polynucleotide is introduced into the cell and does not integrate into the genome of the cell.

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.

"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.

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 plant). 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.

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.

A nucleic acid construct of the invention can be introduced into a cell by any method known to those of skill in the art. In some embodiments of the invention, transformation of a cell comprises nuclear transformation. In other embodiments, transformation of a cell comprises plastid transformation (e.g., chloroplast transformation). In still further embodiments, the recombinant nucleic acid construct of the invention can be introduced into a cell via conventional breeding techniques.

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)).

A nucleotide sequence 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 sequences 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 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, the nucleotide sequences 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.

The present invention is directed to improved base editing nucleic acid constructs. In some embodiments, the present invention provides a nucleic acid construct comprising: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that is capable of binding to the peptide tag; and (c) a guide nucleic acid. In some embodiments, the present invention provides a nucleic acid construct comprising: (a) a Cas12a fusion protein comprising a Cas12a effector protein fused to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and (c) a guide nucleic acid.

In some embodiments, the Cas12a (Cpf1) effector protein may be a LbCpf1 [*Lachnospiraceae* bacterium], AsCpf1 [*Acidaminococcus* sp.], BpCpf1 [*Butyrivibrio proteoclasticus*], CMtCpf1 [*Candidatus Methanoplasma termitum*], EeCpf1 [*Eubacterium eligens*], FnCpf1 (*Francisella novicida* U112), Lb2Cpf1 [*Lachnospiraceae* bacterium], >Lb3Cpf1 [*Lachnospiraceae* bacterium], LiCpf1 [*Leptospira inadai*], MbCpf1 [*Moraxella bovoculi* 237], PbCpf1 [Parcubacteria bacterium GWC2011_GWC2_44_17], PcCpf1 [*Porphyromonas crevioricanis*], PdCpf1 [*Prevotella disiens*], PeCpf1 [*Peregrinibacteria* bacterium GW2011_GWA_33_10], PmCpf1 [*Porphyromonas macacae*], and/or a SsCpf1 [*Smithella* sp. SC_K08D17] (e.g., having a sequence of any one of **SEQ ID NOs:3-22**). In some embodiments, the Cas12a effector protein may be a *Lachnospiraceae* bacterium ND2006 Cas12a (LbCas12a)(LbCpf1) (e.g., having a sequence of any one of **SEQ ID NOs:3** and **9-11**), an *Acidaminococcus sp.* Cpf1 (AsCas12a) (AsCpf1) (e.g., having a sequence of any one of **SEQ ID NO:4**) and/or enAsCas12a (e.g., having a sequence of any one of **SEQ ID NOs:20-22**).

In some embodiments, the present invention provides a nucleic acid construct comprising: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to an affinity polypeptide that is capable of binding to a peptide tag; (b)

a deaminase fusion protein comprising a deaminase fused to the peptide tag; and (c) a guide nucleic acid. In some embodiments, the present invention provides a nucleic acid construct comprising: (a) a Cas12a fusion protein comprising a Cas12a effector protein fused to an affinity polypeptide that binds to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to the peptide tag; and (c) a guide nucleic acid.

In some embodiments, a nucleic acid construct is provided comprising:(a) a Type V CRISPR-Cas effector protein ; (b) a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif; and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif. In some embodiments, a nucleic acid construct is provided comprising:(a) a Cas12a effector protein ; (b) a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif; and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif.

In some embodiments, a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system is provided comprising: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and (c) a guide nucleic acid comprising a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Type V CRISPR-Cas effector protein of the Type V CRISPR-Cas fusion protein and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the Type V CRISPR-Cas fusion protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Type V CRISPR-Cas fusion protein and target nucleic acid by the binding of the affinity polypeptide to the peptide tag that is fused to the Type V CRISPR-Cas fusion protein, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid. In some embodiments, a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system is provided comprising: (a) a Cas12a fusion protein comprising a Cas12a effector protein fused to a peptide tag; (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and (c) a guide nucleic acid comprising a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Cas12a effector protein of the Cas12a fusion protein and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the Cas12a fusion protein to the target nucleic acid, and wherein the deaminase fusion protein is

recruited to the Cas12a fusion protein and target nucleic acid by the binding of the affinity polypeptide to the peptide tag that is fused to the Cas12a fusion protein, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid.

In some embodiments, the present invention provides a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system comprising: (a) a Type V CRISPR-Cas effector protein ; (b) a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif, and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, wherein the recruiting guide nucleic acid comprises a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Type V CRISPR-Cas effector protein and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the Type V CRISPR-Cas effector protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Type V CRISPR-Cas effector protein and target nucleic acid by the binding of the affinity polypeptide to the RNA recruiting motif that is fused to the recruiting guide nucleic acid, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid. In some embodiments, the present invention provides a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system comprising: (a) a Cas12a effector protein; (b) a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif, and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, wherein the recruiting guide nucleic acid comprises a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Cas12a effector protein and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the Cas12a effector protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Cas12a effector protein and target nucleic acid by the binding of the affinity polypeptide to the RNA recruiting motif that is fused to the recruiting guide nucleic acid, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid.

In some embodiments, a nucleic acid construct of the invention (e.g., a polynucleotide encoding a Type V CRISPR-Cas effector protein, a polynucleotide encoding a Type V CRISPR-Cas fusion protein, a polynucleotide encoding a deaminase, a polynucleotide encoding a deaminase fusion protein, a polynucleotide encoding a peptide tag, a polynucleotide encoding an affinity polypeptide, an RNA recruiting motif, a recruiting guide nucleic acid 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.

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:1** or **SEQ ID NO:2**). 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.

In some embodiments, a nucleic acid construct of the invention may further encode one or more polypeptides of interest, optionally wherein the one or more polypeptides of interest may be codon optimized for expression in a plant.

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)), 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, ligase activity, helicase activity, a nuclear localization sequence or activity, 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 Type V CRISPR-Cas effector protein to provide a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein and a polypeptide of interest. In some embodiments, a Type V CRISPR-Cas fusion protein that comprises a Type V CRISPR-Cas effector protein linked to a peptide tag or an affinity polypeptide may also be linked to a polypeptide of interest (e.g., a Type V CRISPR-Cas effector protein may be, for example, linked to both a peptide tag (or an affinity polypeptide) and, for example, a polypeptide of interest, e.g., a UGI). In some embodiments, a polypeptide of interest may be a uracil glycosylase inhibitor (e.g., uracil-DNA glycosylase inhibitor (UGI)).

In some embodiments, a nucleic acid construct of the invention encoding a Type V CRISPR-Cas fusion protein, a deaminase fusion protein and comprising a guide nucleic acid may further encode a polypeptide of interest, optionally wherein the polypeptide of interest may be codon optimized for expression in an organism. In some embodiments, a nucleic acid construct of the invention encoding a Type V CRISPR-Cas effector protein, a deaminase fusion protein and comprising a recruiting guide nucleic acid may further encode a polypeptide of interest, optionally wherein the polypeptide of interest may be codon optimized for expression in an organism (e.g., a plant).

As used herein, a "Type V CRISPR-Cas effector protein" or "Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) effector protein" is a protein or polypeptide or domain thereof of the Type V CRISPR-Cas system 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 Type V CRISPR-Cas effector protein may be an enzyme (e.g., a nuclease, endonuclease, nickase, etc.) or portion thereof and/or may function as an enzyme. In some embodiments, a Type V CRISPR-Cas effector protein refers to a Type V CRISPR-Cas nuclease polypeptide or domain that comprises nuclease activity or in which the nuclease activity has been reduced or eliminated, and/or comprises nickase activity or in which the nickase has been reduced or eliminated, and/or comprises single stranded DNA cleavage activity (ss DNAse activity) or in which the ss DNAse activity has been reduced or eliminated, and/or comprises self-processing RNAse activity or in which the self-processing RNAse activity has been reduced or eliminated. A Type V CRISPR-Cas effector protein

may bind to a target nucleic acid. In some embodiments, a Type V CRISPR-Cas effector protein may be a Cas12 effector protein.

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

A Type V CRISPR-Cas effector protein useful with embodiments of the invention may be any Type V CRISPR-Cas nuclease. A Type V CRISPR-Cas nuclease useful with this invention as an effector protein can include, 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 nuclease. In some embodiments, a Type V CRISPR-Cas nuclease polypeptide or domain useful with embodiments of the invention may be a Cas12a polypeptide or domain. In some embodiments, a Type V CRISPR-Cas effector protein useful with embodiments of the invention may be a nickase, optionally, a Cas12a nickase.

In some embodiments, the Type V CRISPR-Cas effector protein may be a Cas12a effector protein. Cas12a differs in several respects from the more well-known Type II CRISPR Cas9. 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.

A Type V CRISPR-Cas effector protein may be a CRISPR-Cas12a polypeptide or CRISPR-Cas12a domain obtained from any known or later identified Cas12a (previously

known as Cpf1) (see, e.g., U.S. Patent No. 9,790,490, which is incorporated by reference for its disclosures of Cpf1 (Cas12a) sequences). The term "Cas12a", "Cas12a polypeptide" or "Cas12a domain" refers to an RNA-guided polypeptide comprising a Cas12a polypeptide, or a fragment thereof, which comprises the guide nucleic acid binding domain of Cas12a and/or an active, inactive, or partially active DNA cleavage domain of Cas12a and/or the RNA-guided polypeptide may be have nuclease activity. 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 domain or Cas12a polypeptide having a mutation in its nuclease active site, and therefore, no longer comprising nuclease activity, is commonly referred to as deadCas12a (e.g., dCas12a). In some embodiments, a Cas12a domain or Cas12a polypeptide having a mutation in its nuclease active site may have impaired activity (e.g., may have impaired nickase activity).

In some embodiments, a Type V CRISPR-Cas effector protein (e.g., a Cas12a polypeptide) may be optimized for expression in an organism, for example, in an animal, a plant, a fungus, an archaeon, or a bacterium. In some embodiments, a Type V CRISPR-Cas effector protein(e.g., Cas12a polypeptide) may be optimized for expression in a plant.

Any deaminase or domain or polypeptide thereof 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 cystosine 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.

A cytosine deaminase useful with this invention may be any known or later identified cytosine deaminase from any organism (see, e.g., U.S. Patent 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 that can catalyze 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).

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. Patent 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, a cytosine deaminase may be an APOBEC1 deaminase, optionally an APOBEC1 deaminase having the amino acid sequence of **SEQ ID NO:23**. In some embodiments, a cytosine deaminase may be an APOBEC3A deaminase, optionally an APOBEC3A deaminase having the amino acid sequence of **SEQ ID NO:24**. In some embodiments, a cytosine deaminase may be an CDA1 deaminase, optionally a CDA1 having the amino acid sequence of **SEQ ID NO:25**. In some embodiments, a cytosine deaminase may be a FERNY deaminase, optionally a FERNY having the amino acid sequence of **SEQ ID NO:26**. In some embodiments, the cytosine deaminase may be a rAPOBEC1 deaminase, optionally a rAPOBEC1 deaminase having the amino acid sequence of **SEQ ID NO:27**. In some embodiments, the cytosine deaminase may be a hAID deaminase, optionally a hAID having the amino acid sequence of **SEQ ID NO:28** or **SEQ ID NO:29**.

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:30**, **SEQ ID NO:31**, **SEQ ID NO:32**). 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 NO:23-32** (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:23-32**). In some embodiments, a polynucleotide encoding a cytosine deaminase may be codon optimized for expression in an organism (e.g., a plant) and the codon optimized polypeptide may be about 70% to 99.5% identical to the reference polynucleotide.

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. Patent No. 10,113,163, which is incorporated by reference herein for its disclosure of adenine deaminases).

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 may be about 70% to 99.9% identical to a naturally occurring adenine deaminase polypeptide (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 may be codon optimized for expression in a plant.

In some embodiments, an adenine deaminase may be a wild type tRNA-specific adenosine deaminase, e.g., a tRNA-specific adenosine deaminase (TadA) and/or a mutated/evolved adenosine deaminase, e.g., mutated/evolved tRNA-specific adenosine deaminase (TadA*). In some embodiments, TadA 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:33**. In some embodiments, a mutated/evolved *E. coli* TadA* comprises the amino acid sequence of **SEQ ID NOs:34-37** (e.g., **SEQ ID NOs: 34, 35, 36,** or **37**). 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:33-43**.

A "uracil glycosylase inhibitor" or "UGI" useful with the invention may be any protein or polypeptide or domain thereof that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI comprises a wild type UGI or a fragment thereof. In some embodiments, a UGI 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 an amino acid sequence of a naturally occurring UGI. In some embodiments, a UGI may comprise the amino acid sequence of **SEQ ID NO:44** or a polypeptide having about 70% to

about 99.5% identity to the amino acid sequence of **SEQ ID NO:44** (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:44**). For example, in some embodiments, a UGI may comprise a fragment of the amino acid sequence of **SEQ ID NO:44** 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:44**. In some embodiments, a UGI may be a variant of a known UGI (e.g., **SEQ ID NO:44**) 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.

In some embodiments, a nucleic acid construct of this invention comprising, for example, a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag, a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag, and a guide nucleic acid; or a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein, a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif and a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif may further comprise/encode a polypeptide of interest. In some embodiments, a polypeptide of interest may be a uracil glycosylase inhibitor (UGI) (e.g., uracil-DNA glycosylase inhibitor) polypeptide or domain, optionally wherein the UGI may be codon optimized for expression in an organism (e.g., a plant). In some embodiments, a glycosylase inhibitor may be fused to a Type V CRISPR-Cas effector protein and/or a deaminase. In some embodiments, a glycosylase inhibitor may be recruited to the Type V CRISPR-Cas effector protein and/or the deaminase via methods and constructs disclosed herein for protein-protein recruitment, protein-RNA recruitment, and/or chemical recruitment. Thus, as an example a glycosylase inhibitor may be recruited to the Type V CRISPR-Cas effector protein and/or deaminase utilizing a peptide tag/affinity polypeptide, an RNA recruiting motif/affinity polypeptide and/or biotin-streptavidin interaction (or other chemical interaction) as described herein.

The nucleic acid constructs of the invention comprising a Type V CRISPR-Cas effector protein or a fusion protein thereof may be used in combination with a guide nucleic acid (e.g., gRNA, CRISPR array, CRISPR RNA, crRNA) or recruiting guide nucleic acid that is designed to function with the encoded Type V CRISPR-Cas effector protein to modify a target nucleic acid. A guide nucleic acid and/or recruiting guide nucleic acid useful with this invention may comprise at least one spacer sequence and at least one repeat sequence. The guide nucleic acids and recruiting guide nucleic acids are capable of forming a complex with a Type V CRISPR-Cas effector protein of the present invention (e.g., a Type V CRISPR-Cas effector protein that is encoded and expressed by a nucleic acid construct of the invention) and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the complex (e.g., the Type V CRISPR-Cas effector protein to the target nucleic acid), whereby the target nucleic acid may be modified (e.g., cleaved or edited) and/or modulated (e.g., modulating transcription) optionally by a deaminase (e.g., a cytosine deaminase and/or adenine deaminase that is optionally present in and/or recruited to the complex).

In some embodiments, a nucleic acid construct encoding a Type V CRISPR-Cas effector protein (e.g., Cas12a, Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c1, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, and/or Cas14c) fused to a peptide tag (e.g., a Type V CRISPR-Cas effector fusion protein), and a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag may be used in combination with a Type V CRISPR-Cas guide nucleic acid to modify a target nucleic acid, wherein the guide nucleic acid binds to the target nucleic acid and guides the Type V CRISPR-Cas effector protein to the target nucleic acid, and the deaminase fusion protein is recruited to the Type V CRISPR-Cas effector protein and to the target nucleic acid via the binding of the affinity polypeptide of the deaminase to the peptide tag of the Type V CRISPR-Cas effector protein, whereby the deaminase of the deaminase fusion protein can then deaminate a cytosine base in the target nucleic acid, thereby modifying (e.g., editing) the target nucleic acid.

In some embodiments, a nucleic acid construct encoding a Type V CRISPR-Cas effector protein (e.g., Cas12a, Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c1, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, and/or Cas14c), and a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif may be used in combination with a recruiting guide nucleic acid comprising a guide nucleic acid linked to an RNA recruiting motif to modify a target nucleic acid, wherein the recruiting guide nucleic acid binds to the

target nucleic acid and guides the Type V CRISPR-Cas effector protein to the target nucleic acid, and the deaminase fusion protein is recruited to the target nucleic acid via the binding of the affinity polypeptide to the RNA recruiting motif of the recruiting guide nucleic acid, whereby the deaminase of the deaminase fusion protein can then deaminate a cytosine base in the target nucleic acid, thereby modifying (e.g., editing) the target nucleic acid.

A "guide nucleic acid," "guide RNA," "gRNA," "CRISPR RNA/DNA" "crRNA", or "crDNA" as used herein refers to a nucleic acid that comprises at least one spacer sequence, which is complementary to (and hybridizes to) a target DNA (e.g., protospacer), and at least one repeat sequence (e.g., a repeat of a Type V CRISPR-Cas system, or a fragment or portion thereof, including but not limited to, Cas12a, Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c1, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, and/or Cas14c, or a fragment thereof), wherein the repeat sequence may be linked to the 5' end and/or the 3' end of the spacer sequence. The design of a gRNA of this invention is based on a Type V CRISPR-Cas system. In some embodiments, the guide nucleic acid comprises DNA. In some embodiments, the guide nucleic acid comprises RNA. In some embodiments, a Type V CRISPR-Cas effector protein, e.g., a Cas12a gRNA, may comprise, from 5' to 3', a repeat sequence (e.g., a full length repeat sequence or portion thereof ("handle"); e.g., pseudoknot-like structure) and a spacer sequence.

A "recruiting guide nucleic acid" or “recruiting guide RNA” as used herein refer to a guide nucleic acid as defined herein that comprises an RNA recruiting motif. In some embodiments, the RNA recruiting motif may be linked to the 3' end or the 5' end of the recruiting guide nucleic acid. In some embodiments, the RNA recruiting motif may be inserted into the recruiting guide nucleic acid (e.g., within the hairpin loop). An RNA recruiting motif useful with this invention may be any RNA motif capable of being recognized by an affinity polypeptide, e.g., the RNA recruiting motif is capable of being bound by the affinity polypeptide. An RNA recruiting motif and its corresponding affinity polypeptide may include, but is not limited to, a telomerase Ku binding motif (e.g., Ku binding hairpin) and an affinity polypeptide of Ku (e.g., Ku heterodimer); a telomerase Sm7 binding motif and an affinity polypeptide of Sm7; an MS2 phage operator stem-loop and an affinity polypeptide of MS2 Coat Protein (MCP), a PP7 phage operator stem-loop and an affinity polypeptide of PP7 Coat Protein (PCP); an SfMu phage Com stem-loop and an affinity polypeptide of Com RNA binding protein; a PUF binding site (PBS) and a corresponding Pumilio/fem-3 mRNA binding factor (PUF); and/or a synthetic RNA-aptamer and a corresponding aptamer ligand. See, for example, WO2018/129129 and U.S. Patent

Application Publication Nos. 20190218261 and 20180094257, each of which is incorporated by reference herein for their disclosure of RNA recruiting motifs.

In some embodiments, a recruiting guide nucleic acid may be linked to one RNA recruiting motif or to two or more RNA recruiting motifs (e.g., about 2, 3, 4, 5, 6, 7, 8, 9, 10 or more copies of an RNA recruiting motif; e.g., about 2 to about 5, about 2 to about 8, about 2 to about 10, about 3 to about 5, about 3 to about 8, about 5 to about 8, about 5 to about 10, and the like, recruiting motifs), optionally wherein the two or more RNA recruiting motifs may be the same RNA recruiting motif or may be different RNA recruiting motifs. 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:45-55**.

In some embodiments, a guide nucleic acid and/or recruiting 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 or recruiting 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, e.g., a recruiting guide nucleic acid.

A "repeat sequence" as used herein, refers to, for example, any repeat sequence of a wild-type Type V CRISPR Cas locus (e.g., Cas12a locus, Cas12b locus, Cas12c locus (C2c3), Cas12d locus (CasY), Cas12e locus (CasX), Cas12g locus, Cas12h locus, Cas12i locus, C2c1 locus, C2c4 locus, C2c5 locus, C2c8 locus, C2c9 locus, C2c10 locus, Cas14a locus, Cas14b locus, and/or Cas14c locus, or a fragment thereof) or a repeat sequence of a synthetic crRNA that is functional with a Type V 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 Type V CRISPR-Cas locus or it can be a synthetic repeat designed to function in a Type V 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 V 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, may be

linked at its 3' end to the 5' end of a spacer sequence, thereby forming a repeat-spacer sequence (e.g., guide RNA, crRNA).

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 10 to about 100, about 15 to about 30, about 15 to about 40, about 15 to about 45, about 15 to about 50, about 50 to about 100, about 20 to about 30, about 20 to about 40, about 20 to about 50, about 20 to about 100, about 30 to about 40, about 30 to about 50, about 30 to about 100, about 40 to about 80, about 40 to about 100, about 50 to about 100 or more nucleotides.

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").

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.

In some embodiments, the 5' region of a spacer sequence of a guide nucleic acid may be identical to a target DNA, while the 3' region of the spacer may be substantially complementary to the target DNA (e.g., Type V CRISPR-Cas) or the 3' region of a spacer sequence of a guide nucleic acid may be identical to a target DNA, and therefore, the overall complementarity of the spacer sequence to the target DNA may be less than 100%. Thus, for example, in a guide for a Type V CRISPR-Cas system, the first 1, 2, 3, 4, 5, 6, 7, 8, 9, or 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 DNA, 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 DNA. 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 DNA, 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 DNA. A recruiting guide nucleic acid further comprises one or more recruiting motifs as described herein, which may be linked to the 5' end of the guide nucleic acid, the 3' end of the guide nucleic acid or the one or more RNA recruiting motif(s) may be inserted into the recruiting guide nucleic acid (e.g., within the hairpin loop).

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.

As used herein, a "target nucleic acid", "target DNA," "target nucleotide sequence," "target region," or "target region in the genome" refer to a region of an organism's genome

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 of this invention. A target region useful for a Type V CRISPR-Cas system, known as the protospacer adjacent motif (PAM), may be located adjacent to the spacer (or target) sequence. These PAM DNA sequences are typically described by referencing their sequence and location with respect to the non-target strand of the CRISPR complex. These PAM sequences can be either 3' (e.g., Type V CRISPR-Cas system) or 5' (e.g., Type II CRISPR-Cas system) to the ends of a protospacer sequence. A target region (also referred to as the protospacer) 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 or more nucleotides, and the like) located immediately adjacent to a PAM sequence.

A "protospacer sequence" refers to the target double stranded DNA and specifically to the portion of the target DNA (e.g., or target region in the genome) 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).

In the case of Type V CRISPR-Cas (e.g., Cas12a) systems, the protospacer sequence is flanked (e.g., immediately adjacent to) by a protospacer adjacent motif (PAM). For Type V 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'-NNNNNNNNNNNNNNNNNNN-3' RNA Spacer **(SEQ ID NO:56)**
| | | | | | | | | | | | | | | | | | | |
3'**AAAN**NNNNNNNNNNNNNNNNNN-5' Target strand **(SEQ ID NO:57)**
| | | |
5'**TTTN**NNNNNNNNNNNNNNNNNN-3' Non-target strand **(SEQ ID NO:58)**

Guide structures and PAMs are described in by R. Barrangou (*Genome Biol.* 16:247 (2015)).

Canonical Type V CRISPR-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, non-canonical PAMs may be used but may be less efficient.

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).

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 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:59** and **SEQ ID NO:60**. An affinity polypeptide useful with a peptide tag includes, but is not limited to, **SEQ ID NO:61**.

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., peptide 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 peptide tag(s)). 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. Patent 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.

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 nucleic acid binds to the target nucleic acid and the RNA recruiting motif binds to the affinity polypeptide, thereby recruiting the polypeptide to the guide nucleic acid 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).

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; and/or a bifunctional ligand (e.g., fusion of two protein-binding chemicals together; e.g. dihyrofolate reductase (DHFR).

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 Type V CRISPR-Cas fusion protein of the invention may comprise a Type V 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).

In some embodiments, a peptide tag may be fused to a Type V CRISPR-Cas protein. In some embodiments, a peptide tag may be fused or linked to the C-terminus of a Type V CRISPR-Cas effector protein to form a Type V CRISPR-Cas fusion protein. In some embodiments, a peptide tag may be fused or linked to the N-terminus of a Type V CRISPR-Cas effector protein to form a Type V CRISPR-Cas fusion protein.

In some embodiments, when a peptide tag comprises more than one peptide tag, the quantity and/or spacing of an epitope may be optimized within the peptide tag to maximize occupation of the peptide tags and minimize steric interference of, for example, deaminase domains, with each other.

An "affinity polypeptide" (e.g., "recruiting polypeptide") refers to any polypeptide that is capable of binding to its corresponding 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. 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), optionally to recruit the deaminase to a recruiting guide nucleic acid or Type V CRISPR-Cas effector protein. In some embodiments, the affinity polypeptide is stable under the reducing conditions of a cell or cellular extract.

The nucleic acid constructs of the invention and/ guide nucleic acids and/or recruiting 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 a recruiting guide nucleic acid.

When used in combination with guide nucleic acids and recruiting guide nucleic acids, the nucleic acid constructs of the invention (and expression cassettes and vectors 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.

The present invention further provides a method for modifying a target nucleic acid using a composition, complex (e.g., a ribonucleocomplex), system, nucleic acid construct, expression cassette and/or vector of the present invention. The methods may be carried out in an *in vivo* system (e.g., in a cell or in an organism) or in an *in vitro* system (e.g., cell free). In some embodiments, the invention provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas effector protein; (b) a deaminase, optionally wherein the target nucleic acid is contacted with two or more deaminases; and(c) a guide nucleic acid, wherein the deaminase is recruited to the Type V CRISPR-Cas effector protein (e.g., recruited via a protein to protein interaction,

RNA to protein interaction, and/or chemical interaction), optionally wherein the Type V CRISPR-Cas effector protein, the deaminase and guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

In some embodiments, a method of modifying a target nucleic acid is provided, the method comprising: contacting the target nucleic acid with: (a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag (e.g., an epitope or a multimerized epitope); (b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag, optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and (c) a guide nucleic acid, optionally wherein the Type V CRISPR-Cas fusion protein, the deaminase fusion protein and guide nucleic acid are co-expressed, thereby modifying the target nucleic acid. In some embodiments, the peptide tag may be one copy of a peptide tag, or two or more copies (e.g., two or more epitopes) of a peptide tag as described herein. In some embodiments, the peptide tag may be, for example, a GCN4 peptide tag (e.g., Sun-Tag) comprising one to about twenty five repeat units. In some embodiments, the affinity polypeptide may be an antibody. In some embodiments, multiple deaminases may be contacted with the target nucleic acid and recruited by the Type V CRISPR-Case fusion protein. In some embodiments, the target nucleic acid may be contacted with more than one guide nucleic acid, which may comprise the same or different spacer and/or repeat as one another, thereby allowing the targeting of different sites on the target nucleic acid and/or interacting with different Type V CRISPR-Cas effector proteins (e.g., Cas12a, Cas12b, Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), Cas12g, Cas12h, Cas12i, C2c1, C2c4, C2c5, C2c8, C2c9, C2c10, Cas14a, Cas14b, and/or Cas14c).

In some embodiments, the invention provides a method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with: (a) a Type V CRISPR-Cas effector protein; (b) a recruiting guide nucleic acid comprising a guide RNA linked to an RNA recruiting motif, and (c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, optionally wherein the target nucleic acid may be contacted with two or more deaminase fusion proteins, wherein the Type V CRISPR-Cas effector protein, the deaminase fusion protein and the recruiting guide nucleic acid are co-expressed, thereby modifying the target nucleic acid. Any RNA recruiting motif(s) as described herein may be used with the methods of this invention. In some embodiments, multiple deaminases may be contacted with the target nucleic acid and may be recruited by the recruiting guide nucleic acid. In some embodiments, the target

nucleic acid may be contacted with more than one recruiting guide nucleic acid, which may comprise the same or different RNA recruiting motif and/or spacer and/or repeat as one another, thereby allowing the targeting of different sites (e.g., 2, 3, 4, 5, or more different sites) on the target nucleic acid, allowing interaction with different Type V CRISPR-Cas effector proteins and recruiting of multiple polypeptides, which polypeptides may be the same or different.

In some embodiments, a deaminase useful for modifying a target nucleic acid may be a cytosine deaminase and/or an adenine deaminase as described herein. In some embodiments, the cytosine deaminase may be an apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) domain, a human activation induced deaminase (hAID), a FERNY deaminase, and/or a CDA1 deaminase. In some embodiments, the APOBEC deaminase may be an APOBEC3A deaminase. In some embodiments, the adenine deaminase may be a TadA (tRNA-specific adenosine deaminase) and/or TadA* (evolved tRNA-specific adenosine deaminase). In some embodiments, the methods of the invention may further comprise introducing/expressing a glycosylase inhibitor and/or a polynucleotide encoding a glycosylase inhibitor (e.g., a uracil-DNA glycosylase inhibitor (UGI)), optionally wherein method may comprise introducing or expressing two or more glycosylase inhibitors. In some embodiments, the glycosylase inhibitor may be fused to the Type V CRISPR-Cas effector protein and/or the deaminase. In some embodiments, a glycosylase inhibitor may be recruited to a Type V CRISPR-Cas effector protein and/or a deaminase via methods and constructs disclosed herein for protein-protein recruitment, protein-RNA recruitment, and/or chemical recruitment. Thus, as an example a glycosylase inhibitor may be recruited to a Type V CRISPR-Cas effector protein and/or deaminase utilizing a peptide tag/affinity polypeptide, an RNA recruiting motif/affinity polypeptide and/or biotin-streptavidin interaction (or other chemical interaction) as described herein.

The methods of the invention may comprise contacting a target nucleic acid with CRISPR Cas effector proteins, deaminases, and/or fusion proteins thereof of the invention and/or polypeptides of interest, or the target nucleic acid may be contacted with polynucleotides encoding the CRISPR Cas effector proteins, deaminases, and fusion proteins thereof of the invention and/or polypeptides of interest, which polypeptides may be optionally comprised in one or more expression cassettes and/or vectors as described herein, said expression cassettes and/or vectors optionally comprising one or more guide nucleic acids/recruiting guide nucleic acids.

As described herein, the nucleic acids of the invention and/or expression cassettes and/or vectors comprising the same 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, any plant, any fungus, any archaeon, or any bacterium. In some embodiments, the organism may be a plant or cell thereof.

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 the nucleic acid constructs of this 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.

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.

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.

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.

In some embodiments, the invention provides a kit for comprising one or more nucleic acid constructs of the 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 and/or recruiting guide nucleic acid (corresponding to the CRISPR-Cas effector protein encoded by the polynucleotide of the invention) and/or expression cassettes and/or vectors and or cells comprising the same. In some embodiments, a guide nucleic acid and/or recruiting 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 and/or recruiting 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.

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 and/or recruiting 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 and/or recruiting guide nucleic acid.

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).

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. *SunTag-fused dCpf1 for C to T base editing*

Eight copies of GCN4 epitopes were fused to the C-terminus of catalytically deactivated LbCpf1 (dLbCpf1)(dLbCas12a) such that a sequence of **SEQ ID NO:62** was provided. In a separate plasmid, an antibody targeted toward the GCN4 epitope was fused to a variety of deaminases (scFv-deaminase) including rAPOBEC1, hAPOBEC3A, hAPOBEC3B, hAID, and pmCDA1 (**SEQ ID NOs:63-67**).

Plasmids encoding dLbCpf1-SunTag, scFv-deaminase, UGI, and a guide RNA were transfected in HEK293T cells. UGI was co-expressed to transiently suppress base excision repair during the base editing event, which improves efficiency and product purity (Nishida et al. *Science* 353(6305) (2016) (*DOI: 10.1126/science.aaf8729*)). A total of four different guide RNAs targeting endogenous genes were used to probe C to T base editing (**Table 1**).

After three days, genome editing was quantified by next generation sequencing (NGS). While most deaminases tested showed varied and generally low C to T editing efficiencies, APOBEC3A showed consistently high C to T editing rates, ranging from ~7% to ~19% of treated cell population (not sorted for transfected cells) (**Figs. 1-4**). SunTag fusion to Cpf1 is also functional in recruiting APOBEC3A since SunTag fusion provides up to 5 times higher editing efficiency compared to unfused dCpf1 (**Figs. 1-4**). Efficient editing is observed roughly within position 6-15 (position -4,-3,-2,-1 is designated as the PAM (TTTV)) (**Figs. 1-4**).

Table 1: Targeted loci and corresponding spacer sequence.

Loci Name	Guide Name	Spacer Sequence
FANCF-002	PWg120029	ACCTTGGAGACGGCGACTCTCTG (**SEQ ID NO:68**)
FANCF-002	PWg120360	GCGGATGTTCCAATCAGTACGCA (**SEQ ID NO:69**)
AAVS1	PWg120300	TGTCACCAATCCTGTCCCTAGTG (**SEQ ID NO:70**)
AAVS1	PWg120301	TCTGTCCCCTCCACCCCACAGTG (**SEQ ID NO:71**)

Figs. 1-4 show that SunTag-fused dCpf1 can efficiently recruit deaminase domains that are co-expressed in the cell, and this can be used to effect significant levels of C to T base editing. This is the first demonstration of base editing using Cpf1 without the use of the fusion architecture. Multiple additional components as described herein may be co-expressed either singly, expressed under single promoter or any combination.

Example 2. ***SunTag-fused dCpf1 for A to G base editing***

Eight copies of GCN4 epitopes were fused to the C-terminus of catalytically deactivated LbCpf1 (dLbCpf1)(dLbCas12a) such that a sequence of **SEQ ID NO:62** was provided. In a separate plasmid, an antibody targeted toward the GCN4 epitope was fused to an evolved adenine deaminase (TadA8e) (scFv-deaminase) such that a sequence of **SEQ ID NO:72** was provided.

Plasmids encoding dLbCpf1-SunTag, scFv-TadA8e (Richter et al. Nat Biotechnol. 2020, 38(7):883-891) and a guide RNA were transfected in HEK293T cells. A total of five different guide RNAs including a spacer sequence of one of **SEQ ID NOs:73-77** were used to target endogenous genes (Sites 1-5) to probe A to G base editing (**Fig. 5**). After three days,

genome editing is quantified by next generation sequencing (NGS).

Significant adenine deamination was observed at the expected targeting window around position 8-10 (position -4,-3,-2,-1 is designated as the PAM (TTTV)).

The constructs and methods of this invention are broadly applicable to many different types of systems including *in vit*ro and i*n vivo* systems, and may utilize multiple different Type V CRISPR Cas effector proteins and multiple different deaminases, and in multiple types of organisms including animal (e.g., mammalian) and plant systems. By multiplexing guides, editing could be targeted toward multiple different loci within the genome at the same time.

Example 3. SunTag-fused dCpf1 for C to T base editing in soybean plant

T-DNA vectors were constructed that contained expression cassettes for GCN4 epitope-tagged dCpf1 (LbCpf1 and EnAsCpf1), a single-chain antibody-fused APOBEC3A (scFv-A3A), and a uracil glycosylase inhibitor (UGI). These components were either expressed via a single promoter, utilizing fusion and P2A linkers, or via multiple promoters driving expression of individual components (**Table 2**). Specifically, DaMV promoter was used to express the CRISPR-SunTag component, and Mt.Ubq2 promoter was used to express the deaminase and UGI components (**Table 2**). The TDNA vectors also included an antibiotic selection cassette and a guide RNA cassette, driven by a U6 promoter, including a sequence targeting either Target #1 or Target #2 in the target soybean gene. The T-DNA vectors were transformed into Agrobacteria and treated with soybean dried excised embryos to induce plant transformation. The leaves grown in antibiotic selection media were sampled around 4 weeks post-transformation. After extracting DNA, their genetic composition was analyzed using Illumina high-throughput amplicon sequencing.

Base editing activity (C to T change in the target spacer sequence) was detected in all the constructs tested at a rate varying from 20% to 88% (**Table 2**). Among the edited plants for each construct, the high-throughput sequencing analysis showed that each plant sample contained 0.77%-17.48% edited DNA per plant on average (**Table 2**). Editing was observed for both target genes tested (Target #1 and Target #2), and for both CRISPR enzymes tested (LbCpf1 and EnAsCpf1) (**Table 2**). These results demonstrate that this approach for base editing is suitable for other target genes and other Type V CRISPR systems.

Table 2: Expression constructs and their editing results.

Expression constructs used	Target Gene	# plants assayed	# Edited plants	Global edit %	% Average editing per edited plant	% Standard Deviation per edited plant
DaMV promoter: dLbCas12a-8xGCN4, MtUbq2 promoter: scFv-A3A-UGI-GB1	**#1**	**94**	**83**	**88%**	**6.91**	**11.77**
DaMV promoter: dEnAsCas12a-8xGCN4, MtUbq2 promoter: scFv-A3A-GB1-P2A-UGI	**#1**	**94**	**41**	**44%**	**6.58**	**13.81**
MtUbq2 promoter: dLbCas12a-8xGCN4-P2A-scFv-A3A-UGI-GB1	**#1**	**94**	**30**	**32%**	**1.88**	**10.92**
DaMV promoter: dEnAsCas12a-8xGCN4, MtUbq2 promoter: scFv-A3A-UGI-GB1	**#1**	**93**	**19**	**20%**	**0.77**	**0.79**
DaMV promoter: dEnAsCas12a-8xGCN4, MtUbq2 promoter: scFv-A3A-GB1, 35S promoter: UGI	**#1**	**83**	**38**	**46%**	**4.56**	**5.99**
DaMV promoter: dLbCas12a-8xGCN4, MtUbq2 promoter: scFv-A3A-UGI-GB1	**#2**	**94**	**44**	**47%**	**8.53**	**11.21**
DaMV promoter: dLbCas12a-8xGCN4, MtUbq2 promoter: scFv-A3A-GB1-P2A-UGI	**#2**	**88**	**57**	**65%**	**17.48**	**21.14**

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.

THAT WHICH IS CLAIMED IS:

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

contacting the target nucleic acid with:

(a) a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) effector protein;

(b) a deaminase, optionally wherein the target nucleic acid is contacted with two or more deaminases; and

(c) a guide nucleic acid, wherein the deaminase is recruited to the Type V CRISPR-Cas effector protein (e.g., recruited via a protein to protein interaction, RNA to protein interaction, and/or chemical interaction), thereby modifying the target nucleic acid, optionally wherein the Type V CRISPR-Cas effector protein, the deaminase and guide nucleic acid are co-expressed.

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

contacting the target nucleic acid with:

(a) a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag;

(b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag, optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and

(c) a guide nucleic acid, optionally wherein the Type V CRISPR-Cas fusion protein, the deaminase fusion protein and guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

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

contacting the target nucleic acid with:

(a) a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) fusion protein comprising a Type V CRISPR-Cas effector protein fused to an affinity polypeptide that binds to a peptide tag;

(b) a deaminase fusion protein comprising a deaminase fused to the peptide tag, optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and

(c) a guide nucleic acid, optionally wherein the Type V CRISPR-Cas fusion protein, the deaminase fusion protein and guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

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

contacting the target nucleic acid with:

(a) a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) effector protein;

(b) a recruiting guide nucleic acid comprising a guide RNA linked to an RNA recruiting motif, and

(c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, optionally wherein the target nucleic acid is contacted with two or more deaminase fusion proteins; and

optionally wherein the Type V CRISPR-Cas effector protein, the deaminase fusion protein and the recruiting guide nucleic acid are co-expressed, thereby modifying the target nucleic acid.

5. The method of claim 2 or claim 3, wherein the peptide tag comprises 2 or more copies of the peptide tag (e.g., two or more epitopes).

6. The method of any one of claims 2, 3 or 5, wherein the peptide tag is a GCN4 peptide repeat unit (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.

7. The method of any one of claims 2, 3, 5, or 6, wherein the affinity polypeptide is an antibody.

8. The method of claim 7, wherein the antibody is a scFv antibody.

9. The method of claim 4, wherein the recruiting guide nucleic acid is linked to two or more RNA recruiting motifs, optionally wherein the two or more RNA recruiting motifs are the same RNA recruiting motif or different RNA recruiting motifs.

10. The method of claim 4 or claim 9, wherein the RNA recruiting motif is a telomerase Ku binding motif (e.g., Ku binding hairpin) and the affinity polypeptide is a Ku polypeptide (e.g., Ku heterodimer); the RNA recruiting motif is a telomerase Sm7 binding motif and the affinity polypeptide is a Sm7 polypeptide; the RNA recruiting motif is an MS2 phage operator stem-loop and the affinity polypeptide is a MS2 Coat Protein (MCP); the RNA recruiting motif is a PP7 phage operator stem-loop and the affinity polypeptide is a PP7 Coat Protein (PCP); the RNA recruiting motif is an SfMu phage Com stem-loop and the affinity polypeptide is a Com RNA binding protein; the RNA recruiting motif is a Pumilio/fem-3 mRNA binding factor (PUF) and the affinity polypeptide is a PUF binding site (PBS) polypeptide; and/or the RNA recruiting motif is a synthetic RNA-aptamer and the affinity polypeptide is the corresponding aptamer ligand.

11. The method of any one of the preceding claims, wherein the Type V CRISPR-Cas effector protein comprises a mutation in a nuclease active site.

12. The method of any one of the preceding claims, wherein the deaminase is a cytosine deaminase and/or an adenine deaminase.

13. The method of any one of the preceding claims, wherein the two or more deaminase fusion proteins comprise the same or different deaminases.

14. The method of claim 13, wherein a first deaminase fusion protein of the two or more deaminase fusion proteins comprises a cytosine deaminase and a second deaminase fusion protein of the two or more deaminase fusion proteins comprises an adenine deaminase.

15. The method of any one of claims 12 to 14, wherein the cytosine deaminase is an apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC), a human activation induced deaminase (hAID), a FERNY deaminase, and/or a CDA1 deaminase, optionally wherein the cytosine deaminase comprises or is the sequence of any one of **SEQ ID NOs:23-32**.

16. The method of claim 15, wherein the cytosine deaminase is an APOBEC3A, optionally wherein the APOBEC3A comprises or is the sequence of (e.g., **SEQ ID NO:24**).

17. The method of any one of claims 12 to 16, wherein the adenine deaminase is TadA (tRNA-specific adenosine deaminase) and/or TadA* (evolved tRNA-specific adenosine deaminase), optionally wherein the adenine deaminase comprises or is the sequence of any one of **SEQ ID NOs:33-43**.

18. The method of any one of the preceding claims, further comprising introducing a glycosylase inhibitor, optionally introducing two or more glycosylase inhibitors.

19. The method of claim 18, wherein the glycosylase inhibitor is a uracil-DNA glycosylase inhibitor (UGI), optionally wherein the UGI comprises or is the sequence of **SEQ ID NO:44**.

20. The method claim 18 or claim 19, wherein the glycosylase inhibitor is recruited to the Type V CRISPR-Cas effector protein and/or the deaminase, optionally via a protein-protein interaction, RNA-protein interaction, or chemical interaction.

21. The method of claim 18-20, wherein the glycosylase inhibitor is fused to the Type V CRISPR-Cas effector protein and/or the deaminase.

22. The method of any one of the preceding claims, wherein the Type V CRISPR-Cas effector protein, the Type V CRISPR-Cas fusion protein, the deaminase, and/or the deaminase fusion protein is encoded by a polynucleotide.

23. The method of any one of claims 18 to 22, wherein the glycosylase inhibitor is encoded by a polynucleotide.

24. The method of claim 22 or 23, wherein the polynucleotide encoding the Type V CRISPR-Cas effector protein, the polynucleotide encoding the deaminase, and the guide nucleic acid are comprised in one or more expression cassette(s) and/or vector(s).

25. The method of any one of claims 22-24, wherein the polynucleotide encoding the Type V CRISPR-Cas fusion protein, the polynucleotide encoding the deaminase fusion protein, and the guide nucleic acid are comprised in one or more expression cassette(s) and/or vector(s).

26. The method of any one of claims 22-25, wherein the polynucleotide encoding the Type V CRISPR-Cas effector protein, the polynucleotide encoding the deaminase fusion protein, and the recruiting guide nucleic acid are comprised in one or more expression cassette(s) and/or vector(s).

27. The method of any one of claims 23 to 26, wherein the polynucleotide encoding the glycosylase inhibitor is comprised in an expression cassette or vector, optionally in the same expression cassette or vector comprising one or more of the polynucleotide encoding the Type V CRISPR-Cas effector protein, the polynucleotide encoding the deaminase, and the guide nucleic acid, optionally in the same expression cassette or vector comprising one or more of the polynucleotide encoding the Type V CRISPR-Cas fusion protein, the polynucleotide encoding the deaminase fusion protein, and/or the guide nucleic acid, and/or optionally in the same expression cassette or vector comprising one or more of the polynucleotide encoding the Type V CRISPR-Cas effector protein, the polynucleotide encoding the deaminase fusion protein, and/or the recruiting guide nucleic acid.

28. The method of any one of the preceding claims, wherein the Type V CRISPR-Cas effector protein is linked to a polypeptide of interest.

29. The method of claim 28, wherein the polypeptide of interest comprises at least one polypeptide having nickase activity, recombinase activity, transposase activity, methylase activity, glycosylase (DNA glycosylase) activity, glycosylase inhibitor activity (e.g., uracil-DNA glycosylase inhibitor (UGI)), 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, ligase activity, helicase activity, and/or photolyase activity.

30. The method of claim 28 or claim 29, wherein the polypeptide of interest is encoded by a polynucleotide.

31. The method of any one of the preceding claims, wherein the Type V CRISPR-Cas effector protein is a Cas12a (Cpf1) polypeptide, Cas12b polypeptide, Cas12c (C2c3) polypeptide, Cas12d (CasY) polypeptide, Cas12e (CasX) polypeptide, Cas12g polypeptide, Cas12h polypeptide, Cas12i polypeptide, C2c4 polypeptide, C2c5 polypeptide, C2c8 polypeptide, C2c9 polypeptide, C2c10 polypeptide, Cas14a polypeptide, Cas14b polypeptide, and/or Cas14c polypeptide.

32. The method of any of the preceding claims, wherein the Type V CRISPR-Cas effector protein is a Cas12a (Cpf1) polypeptide, optionally wherein the Cas12a (Cpf1) polypeptide has a sequence of any one of **SEQ ID NOs:3-19** or that is encoded by a polynucleotide having a sequence of any one of **SEQ ID NOs:20-22**, optionally wherein the Cas12a polypeptide is a *Lachnospiraceae* bacterium ND2006 Cas12a (LbCas12a)(LbCpf1) polypeptide (e.g., having a sequence of any one of **SEQ ID NOs:3** or **9-11**), an *Acidaminococcus sp.* Cpf1 (AsCas12a) (AsCpf1) polypeptide(e.g., having a sequence of **SEQ ID NO:4**) and/or enAsCas12a polypeptide (e.g., encoded by a polynucleotide having a sequence of any one of **SEQ ID NOs:20-22**).

33. The method of any one of the preceding claims, wherein the target nucleic acid is in an organism, optionally an animal, a plant, a fungus, an archaeon, or a bacterium.

34. The method of claim 28, wherein the polynucleotide encoding the Type V CRISPR-Cas effector protein, the polynucleotide encoding the Type V CRISPR-Cas fusion protein, the polynucleotide encoding the deaminase, the polynucleotide encoding the deaminase fusion protein, the polynucleotide encoding the glycosylase inhibitor, and/or the polynucleotide encoding the polypeptide of interest are codon optimized for expression in the organism comprising the target nucleic acid.

35. A nucleic acid construct comprising:

(a) a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag;

(b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and

(c) a guide nucleic acid.

36. A nucleic acid construct comprising:

(a) a Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) effector protein;

(b) a recruiting guide nucleic acid comprising a guide RNA linked to an RNA recruiting motif; and

(c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif.

37. A Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system comprising:

(a) a Type V CRISPR-Cas fusion protein comprising a Type V CRISPR-Cas effector protein fused to a peptide tag;

(b) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the peptide tag; and

(c) a guide nucleic acid comprising a spacer sequence and a repeat sequence, wherein the guide nucleic acid is capable of forming a complex with the Type V CRISPR-Cas effector protein of the Type V CRISPR-Cas fusion protein and the spacer sequence is capable of hybridizing to a target nucleic acid, thereby guiding the Type V CRISPR-Cas fusion protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Type V CRISPR-Cas fusion protein and target nucleic acid by the binding of the affinity polypeptide to the peptide tag that is fused to the Type V CRISPR-Cas fusion protein, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid.

38. A Type V Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) (CRISPR-Cas) system comprising:

(a) a Type V CRISPR-Cas effector protein;

(b) a recruiting guide nucleic acid comprising a guide RNA linked to an RNA recruiting motif, and

(c) a deaminase fusion protein comprising a deaminase fused to an affinity polypeptide that binds to the RNA recruiting motif, wherein the recruiting guide nucleic acid comprises a spacer sequence and a repeat sequence, wherein the recruiting guide nucleic acid is capable of forming a complex with the Type V CRISPR-Cas effector protein and the guide

RNA is capable of hybridizing to a target nucleic acid, thereby guiding the Type V CRISPR-Cas effector protein to the target nucleic acid, and wherein the deaminase fusion protein is recruited to the Type V CRISPR-Cas effector protein and target nucleic acid by the binding of the affinity polypeptide to the RNA recruiting motif that is fused to the recruiting guide nucleic acid, whereby the system is capable of modifying (e.g., cleaving or editing) the target nucleic acid.

39. The nucleic construct of claim 35 or claim 36, or the CRISPR-Cas system of claim 37 or claim 38, wherein the Type V CRISPR-Cas effector protein is a Cas12a (Cpf1) polypeptide, Cas12b polypeptide, Cas12c (C2c3) polypeptide, Cas12d (CasY) polypeptide, Cas12e (CasX) polypeptide, Cas12g polypeptide, Cas12h polypeptide, Cas12i polypeptide, C2c4 polypeptide, C2c5 polypeptide, C2c8 polypeptide, C2c9 polypeptide, C2c10 polypeptide, Cas14a polypeptide, Cas14b polypeptide, and/or Cas14c polypeptide.

40. The nucleic acid construct of claim 35 or claim 36, or the CRISPR-Cas system of claim 37 or claim 38, wherein the Type V CRISPR-Cas effector protein is a Cas12a (Cpf1) polypeptide, optionally wherein the Cas12a (Cpf1) polypeptide has a sequence of any one of **SEQ ID NOs:3-19** or that is encoded by a polynucleotide having a sequence of any one of **SEQ ID NOs:20-22**, optionally wherein the Cas12a polypeptide is a *Lachnospiraceae* bacterium ND2006 Cas12a (LbCas12a)(LbCpf1) polypeptide (e.g., having a sequence of any one of **SEQ ID NOs:3** or **9-11**), an *Acidaminococcus sp.* Cpf1 (AsCas12a) (AsCpf1) polypeptide(e.g., having a sequence of **SEQ ID NO:4**) and/or enAsCas12a polypeptide (e.g., encoded by a polynucleotide having a sequence of any one of **SEQ ID NOs:20-22**).

41. An expression cassette comprising the nucleic acid construct of any one of claims 35, 36, 39, or 40.

42. An expression cassette comprising the CRISPR-Cas system of any one of claims 37 to 40.

43. A cell comprising the nucleic acid construct of any one of claims 35, 36, 39, or 40, or the expression cassette of claim 41.

44. A cell comprising the CRISPR-Cas system of any one of claims 37 to 40 or the expression cassette of claim 42.

5

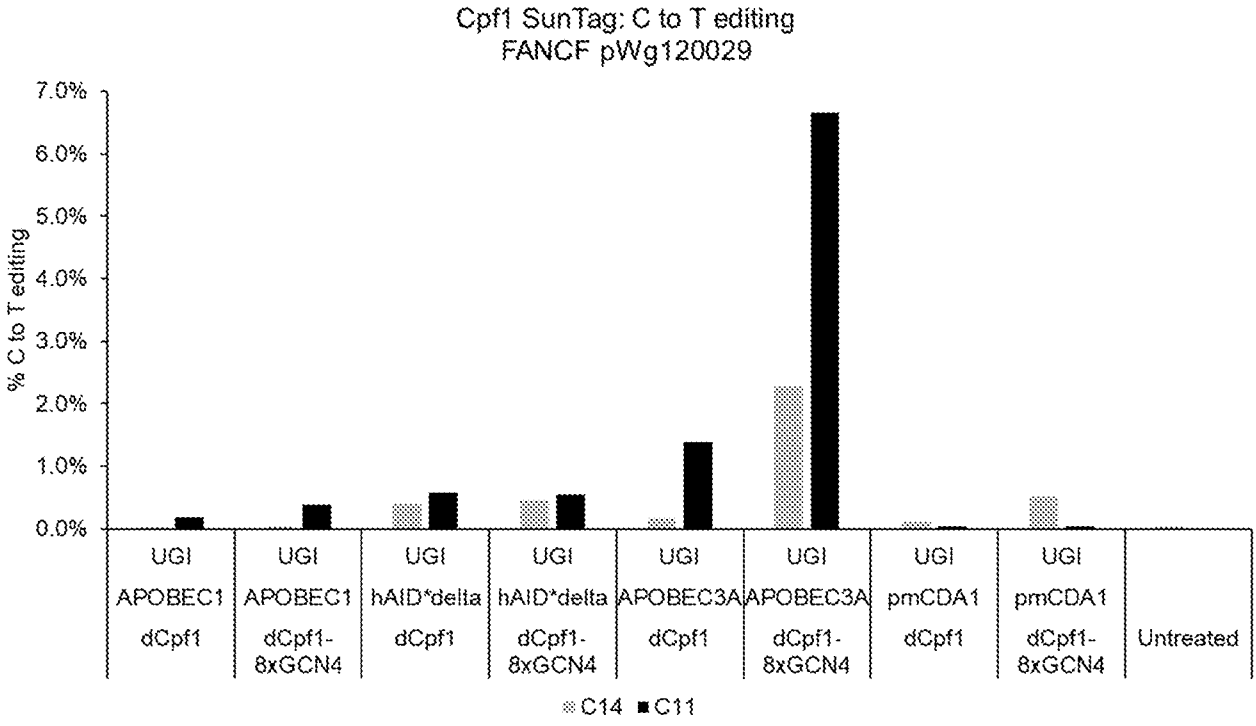

Cpf1 SunTag: C to T editing
FANCF pWg120029
% C to T editing
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
UGI APOBEC1 dCpf1
UGI APOBEC1 dCpf1-8xGCN4
UGI hAID*delta dCpf1
UGI hAID*delta dCpf1-8xGCN4
UGI APOBEC3A dCpf1
UGI APOBEC3A dCpf1-8xGCN4
UGI pmCDA1 dCpf1
UGI pmCDA1 dCpf1-8xGCN4
Untreated
C14
C11

FIG. 1

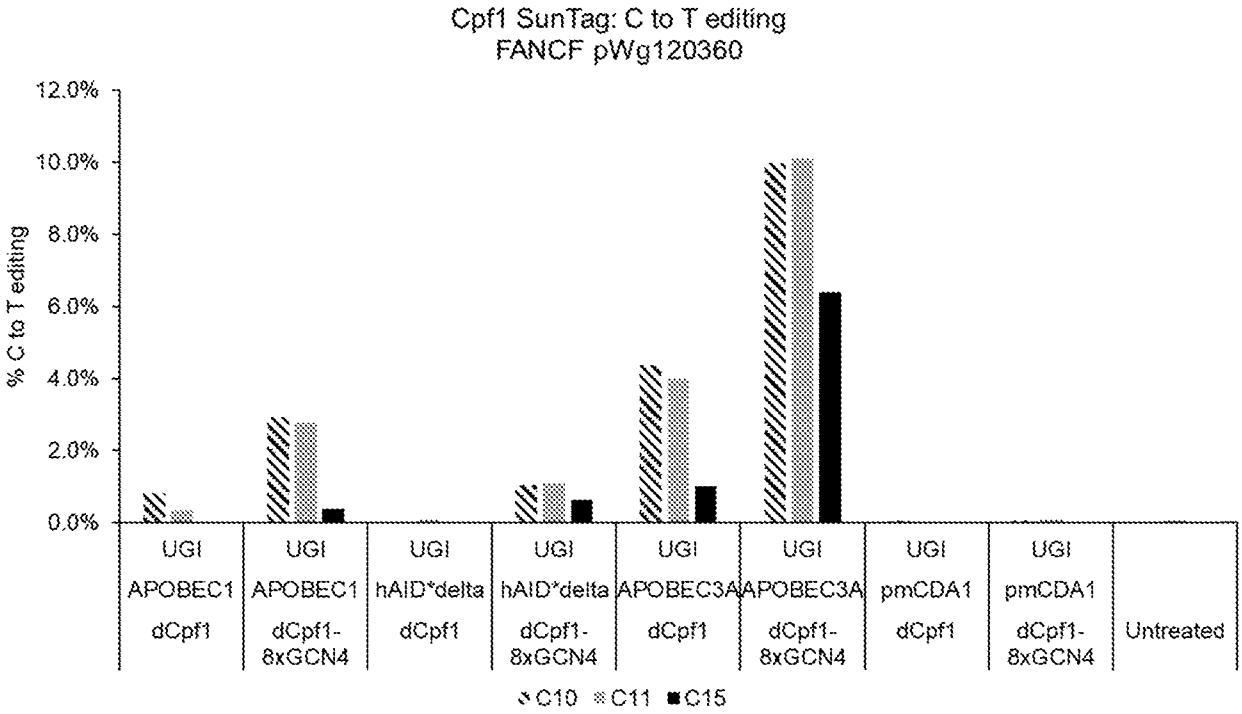

Cpf1 SunTag: C to T editing
FANCF pWg120360
% C to T editing
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
UGI APOBEC1 dCpf1
UGI APOBEC1 dCpf1-8xGCN4
UGI hAID*delta dCpf1
UGI hAID*delta dCpf1-8xGCN4
UGI APOBEC3A dCpf1
UGI APOBEC3A dCpf1-8xGCN4
UGI pmCDA1 dCpf1
UGI pmCDA1 dCpf1-8xGCN4
Untreated
C10
C11
C15

FIG. 2

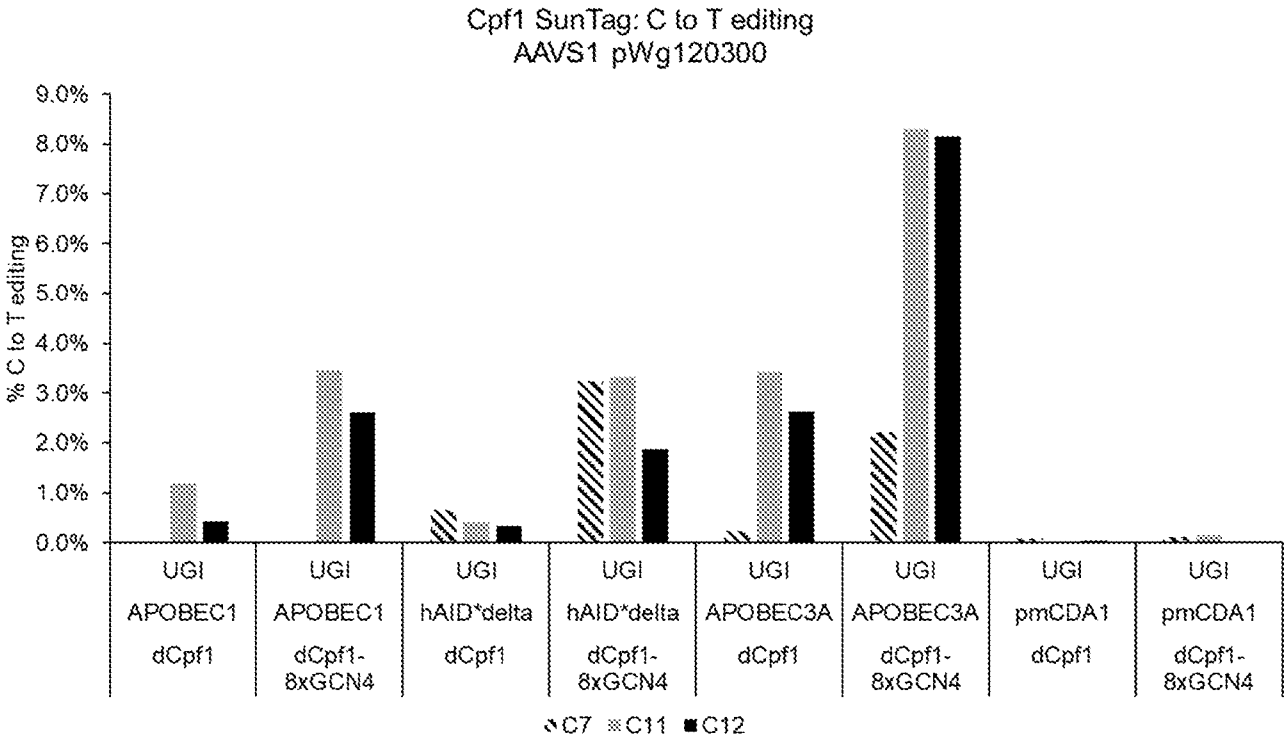

Cpf1 SunTag: C to T editing
AAVS1 pWg120300
% C to T editing
9.0%
8.0%
7.0%
6.0%
5.0%
4.0%
3.0%
2.0%
1.0%
0.0%
UGI APOBEC1 dCpf1
UGI APOBEC1 dCpf1-8xGCN4
UGI hAID*delta dCpf1
UGI hAID*delta dCpf1-8xGCN4
UGI APOBEC3A dCpf1
UGI APOBEC3A dCpf1-8xGCN4
UGI pmCDA1 dCpf1
UGI pmCDA1 dCpf1-8xGCN4
C7 C11 C12

FIG. 3

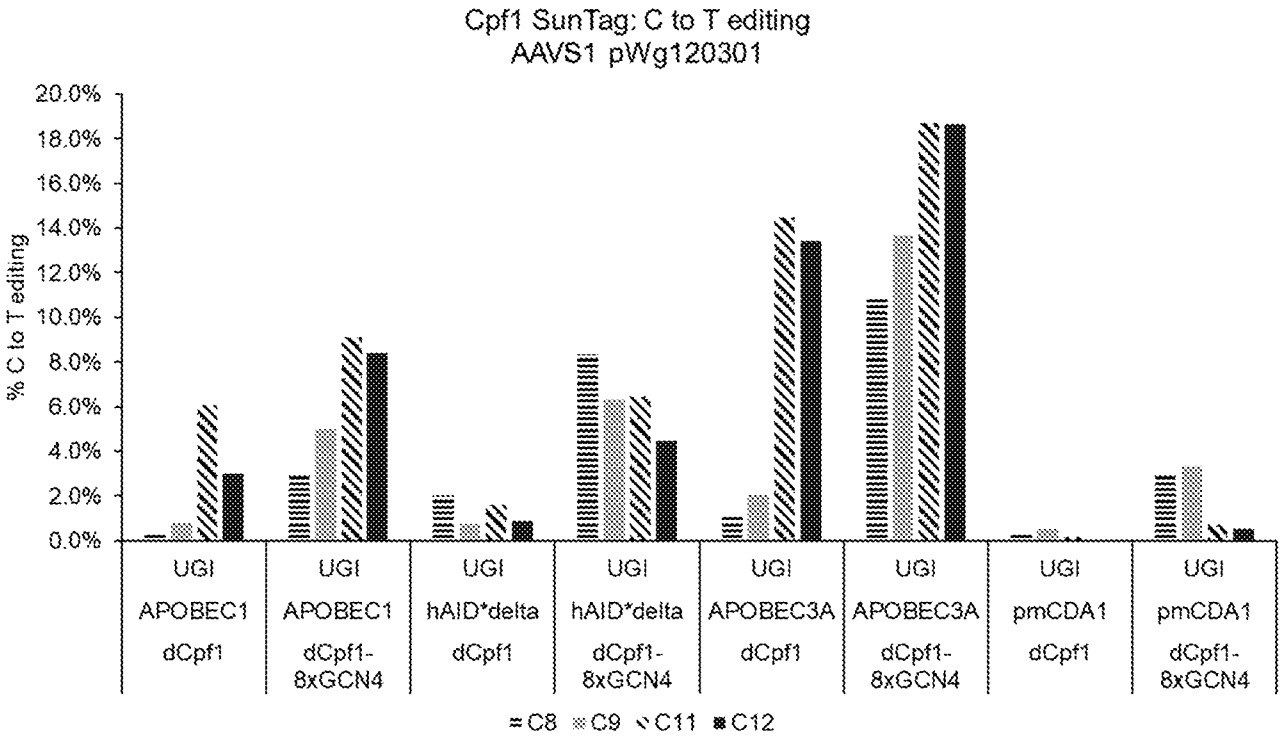

Cpf1 SunTag: C to T editing
AAVS1 pWg120301
% C to T editing
20.0%
18.0%
16.0%
14.0%
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
UGI APOBEC1 dCpf1
UGI APOBEC1 dCpf1-8xGCN4
UGI hAID*delta dCpf1
UGI hAID*delta dCpf1-8xGCN4
UGI APOBEC3A dCpf1
UGI APOBEC3A dCpf1-8xGCN4
UGI pmCDA1 dCpf1
UGI pmCDA1 dCpf1-8xGCN4
C8 C9 C11 C12

FIG. 4

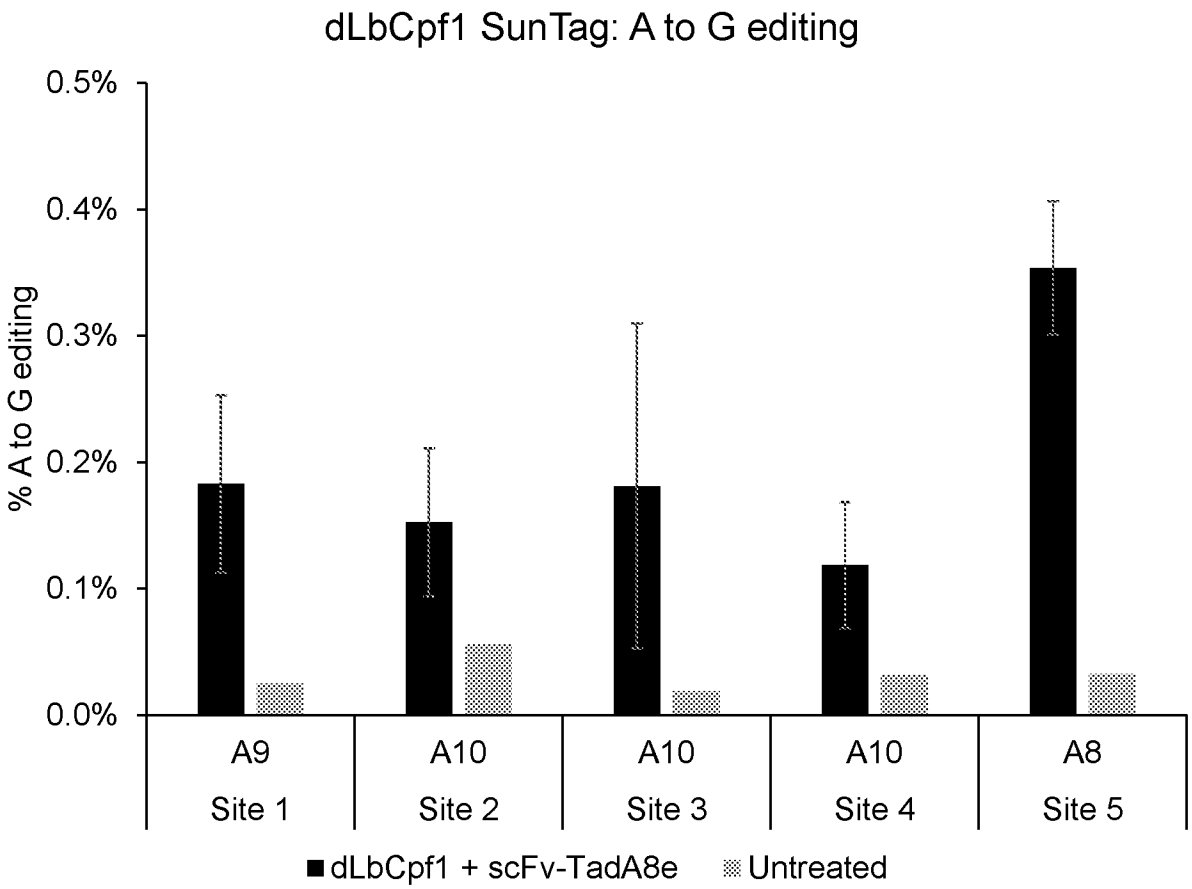

dLbCpf1 SunTag: A to G editing
% A to G editing
0.5%
0.4%
0.3%
0.2%
0.1%
0.0%
A9
Site 1
A10
Site 2
A10
Site 3
A10
Site 4
A8
Site 5
dLbCpf1 + scFv-TadA8e
Untreated

FIG. 5

SEQUENCE LISTING

<110> Pairwise Plants Services, Inc.
Kim, Yongjoo

<120> TYPE V CRISPR-CAS BASE EDITORS AND METHODS OF USE THEREOF

<130> 1499.10.WO

<150> US 62/927,914
<151> 2019-10-30

<160> 77

<170> PatentIn version 3.5

<210> 1
<211> 1592
<212> DNA
<213> Medicago truncatula

<400> 1
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

gtttttttga tgtctgttgt aaattgaatt gattataatt gctgatctgc tgcttccagt 1500

tttcataacc catattcttt taaccttgtt gtacacacaa tgaaaaattg gtgattgatt 1560

catttgtttt tctttgtttt ggattataca gg 1592

<210> 2
<211> 2000
<212> DNA
<213> Zea mays

<400> 2
gtcgtgcccc tctctagaga taaagagcat tgcatgtcta aagtataaaa aattaccaca 60

tatttttttg tcacacttat ttgaagtgta gtttatctat ctctatacat atatttaaac 120

ttcactctac aaataatata gtctataata ctaaaataat attagtgttt tagaggatca 180

tataaataaa ctgctagaca tggtctaaag gataattgaa tattttgaca atctacagtt 240

ttatcttttt agtgtgcatg tgatctctct gtttttttttg caaatagctt gacctatata 300

atacttcatc cattttatta gtacatccat ttaggattta gggttgatgg tttctataga 360

ctaattttta gtacatccat tttattcttt ttagtctcta aatttttttaa aactaaaact 420

ctattttagt tttttattta ataatttaga tataaaatga aataaaataa attgactaca 480

aataaaacaa atacccttta 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 tccccccccc 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

ggatggaaat atcgatctag gataggtata catgttgatg cgggttttac tgatgcatat 1560

acagagatgc ttttttttcgc 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

<210> 3
<211> 1228
<212> PRT
<213> Unknown

<220>
<223> Lachnospiraceae bacterium

<400> 3

Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65 70 75 80

Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
85 90 95

Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
100 105 110

Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
115 120 125

Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
130 135 140

Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145 150 155 160

Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
165 170 175

```
Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
            180                 185                 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
        195                 200                 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
    210                 215                 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225                 230                 235                 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
                245                 250                 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
            260                 265                 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
        275                 280                 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
    290                 295                 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305                 310                 315                 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
                325                 330                 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
            340                 345                 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
        355                 360                 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
    370                 375                 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385                 390                 395                 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu
                405                 410                 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
            420                 425                 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
        435                 440                 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
```

```
    450                 455                 460

Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465                 470                 475                 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
                485                 490                 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
            500                 505                 510

Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
        515                 520                 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
    530                 535                 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545                 550                 555                 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
                565                 570                 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
            580                 585                 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
        595                 600                 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
    610                 615                 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625                 630                 635                 640

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
                645                 650                 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
            660                 665                 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
        675                 680                 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
    690                 695                 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705                 710                 715                 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
                725                 730                 735
```

```
Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
            740                 745                 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
        755                 760                 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
    770                 775                 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785                 790                 795                 800

Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu Val
                805                 810                 815

Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile Ala
            820                 825                 830

Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys Gly
        835                 840                 845

Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe Asn
    850                 855                 860

Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys Glu
865                 870                 875                 880

Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn Ile
                885                 890                 895

Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile Cys
            900                 905                 910

Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu Asn
        915                 920                 925

Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr Gln
    930                 935                 940

Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp Lys
945                 950                 955                 960

Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln Ile
                965                 970                 975

Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly Phe
            980                 985                 990

Ile Phe Tyr Ile Pro Ala Trp Leu  Thr Ser Lys Ile Asp  Pro Ser Thr
        995                 1000                1005
```

```
Gly Phe  Val Asn Leu Leu Lys  Thr Lys Tyr Thr Ser  Ile Ala Asp
    1010                1015                1020

Ser Lys  Lys Phe Ile Ser Ser  Phe Asp Arg Ile Met  Tyr Val Pro
    1025                1030                1035

Glu Glu  Asp Leu Phe Glu Phe  Ala Leu Asp Tyr Lys  Asn Phe Ser
    1040                1045                1050

Arg Thr  Asp Ala Asp Tyr Ile  Lys Lys Trp Lys Leu  Tyr Ser Tyr
    1055                1060                1065

Gly Asn  Arg Ile Arg Ile Phe  Arg Asn Pro Lys Lys  Asn Asn Val
    1070                1075                1080

Phe Asp  Trp Glu Glu Val Cys  Leu Thr Ser Ala Tyr  Lys Glu Leu
    1085                1090                1095

Phe Asn  Lys Tyr Gly Ile Asn  Tyr Gln Gln Gly Asp  Ile Arg Ala
    1100                1105                1110

Leu Leu  Cys Glu Gln Ser Asp  Lys Ala Phe Tyr Ser  Ser Phe Met
    1115                1120                1125

Ala Leu  Met Ser Leu Met Leu  Gln Met Arg Asn Ser  Ile Thr Gly
    1130                1135                1140

Arg Thr  Asp Val Asp Phe Leu  Ile Ser Pro Val Lys  Asn Ser Asp
    1145                1150                1155

Gly Ile  Phe Tyr Asp Ser Arg  Asn Tyr Glu Ala Gln  Glu Asn Ala
    1160                1165                1170

Ile Leu  Pro Lys Asn Ala Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala
    1175                1180                1185

Arg Lys  Val Leu Trp Ala Ile  Gly Gln Phe Lys Lys  Ala Glu Asp
    1190                1195                1200

Glu Lys  Leu Asp Lys Val Lys  Ile Ala Ile Ser Asn  Lys Glu Trp
    1205                1210                1215

Leu Glu  Tyr Ala Gln Thr Ser  Val Lys His
    1220                1225
```

<210> 4
<211> 1307
<212> PRT
<213> Acidaminococcus sp.

<400> 4

```
Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
```

```
1               5                   10                  15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
            20                  25                  30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
        35                  40                  45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
    50                  55                  60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65                  70                  75                  80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
                85                  90                  95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
            100                 105                 110

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
        115                 120                 125

Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
    130                 135                 140

Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
145                 150                 155                 160

Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
                165                 170                 175

Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
            180                 185                 190

Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
        195                 200                 205

Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
    210                 215                 220

Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
225                 230                 235                 240

Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
                245                 250                 255

Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
            260                 265                 270

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
        275                 280                 285
```

```
Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
    290                 295                 300

Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
305                 310                 315                 320

Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
                325                 330                 335

Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
            340                 345                 350

Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
        355                 360                 365

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
    370                 375                 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385                 390                 395                 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
                405                 410                 415

Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
            420                 425                 430

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
        435                 440                 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
    450                 455                 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465                 470                 475                 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
                485                 490                 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
            500                 505                 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
        515                 520                 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
    530                 535                 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545                 550                 555                 560
```

```
Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
                565                 570                 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
            580                 585                 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
        595                 600                 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr
    610                 615                 620

Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys
625                 630                 635                 640

Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln
                645                 650                 655

Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala
            660                 665                 670

Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr
        675                 680                 685

Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr
    690                 695                 700

Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His
705                 710                 715                 720

Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu
                725                 730                 735

Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys
            740                 745                 750

Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu
        755                 760                 765

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
    770                 775                 780

Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
785                 790                 795                 800

Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
                805                 810                 815

Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
            820                 825                 830
```

```
Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
        835                 840                 845

Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
    850                 855                 860

Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
865                 870                 875                 880

Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
                885                 890                 895

Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Asp Arg Gly Glu Arg
            900                 905                 910

Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
        915                 920                 925

Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
    930                 935                 940

Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
945                 950                 955                 960

Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
                965                 970                 975

His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
            980                 985                 990

Glu Asn Leu Asn Phe Gly Phe Lys  Ser Lys Arg Thr Gly  Ile Ala Glu
        995                 1000                 1005

Lys Ala  Val Tyr Gln Gln Phe  Glu Lys Met Leu Ile  Asp Lys Leu
    1010                1015                1020

Asn Cys  Leu Val Leu Lys Asp  Tyr Pro Ala Glu Lys  Val Gly Gly
    1025                1030                1035

Val Leu  Asn Pro Tyr Gln Leu  Thr Asp Gln Phe Thr  Ser Phe Ala
    1040                1045                1050

Lys Met  Gly Thr Gln Ser Gly  Phe Leu Phe Tyr Val  Pro Ala Pro
    1055                1060                1065

Tyr Thr  Ser Lys Ile Asp Pro  Leu Thr Gly Phe Val  Asp Pro Phe
    1070                1075                1080

Val Trp  Lys Thr Ile Lys Asn  His Glu Ser Arg Lys  His Phe Leu
    1085                1090                1095

Glu Gly  Phe Asp Phe Leu His  Tyr Asp Val Lys Thr  Gly Asp Phe
```

```
    1100                1105                1110

Ile Leu  His Phe Lys Met Asn  Arg Asn Leu Ser Phe  Gln Arg Gly
    1115                1120                1125

Leu Pro  Gly Phe Met Pro Ala  Trp Asp Ile Val Phe  Glu Lys Asn
    1130                1135                1140

Glu Thr  Gln Phe Asp Ala Lys  Gly Thr Pro Phe Ile  Ala Gly Lys
    1145                1150                1155

Arg Ile  Val Pro Val Ile Glu  Asn His Arg Phe Thr  Gly Arg Tyr
    1160                1165                1170

Arg Asp  Leu Tyr Pro Ala Asn  Glu Leu Ile Ala Leu  Leu Glu Glu
    1175                1180                1185

Lys Gly  Ile Val Phe Arg Asp  Gly Ser Asn Ile Leu  Pro Lys Leu
    1190                1195                1200

Leu Glu  Asn Asp Asp Ser His  Ala Ile Asp Thr Met  Val Ala Leu
    1205                1210                1215

Ile Arg  Ser Val Leu Gln Met  Arg Asn Ser Asn Ala  Ala Thr Gly
    1220                1225                1230

Glu Asp  Tyr Ile Asn Ser Pro  Val Arg Asp Leu Asn  Gly Val Cys
    1235                1240                1245

Phe Asp  Ser Arg Phe Gln Asn  Pro Glu Trp Pro Met  Asp Ala Asp
    1250                1255                1260

Ala Asn  Gly Ala Tyr His Ile  Ala Leu Lys Gly Gln  Leu Leu Leu
    1265                1270                1275

Asn His  Leu Lys Glu Ser Lys  Asp Leu Lys Leu Gln  Asn Gly Ile
    1280                1285                1290

Ser Asn  Gln Asp Trp Leu Ala  Tyr Ile Gln Glu Leu  Arg Asn
    1295                1300                1305


<210>  5
<211>  1241
<212>  PRT
<213>  Butyrivibrio proteoclasticus

<400>  5

Met Leu Leu Tyr Glu Asn Tyr Thr Lys Arg Asn Gln Ile Thr Lys Ser
1               5                   10                  15

Leu Arg Leu Glu Leu Arg Pro Gln Gly Lys Thr Leu Arg Asn Ile Lys
            20                  25                  30
```

```
Glu Leu Asn Leu Leu Glu Gln Asp Lys Ala Ile Tyr Ala Leu Leu Glu
        35                  40                  45

Arg Leu Lys Pro Val Ile Asp Glu Gly Ile Lys Asp Ile Ala Arg Asp
    50                  55                  60

Thr Leu Lys Asn Cys Glu Leu Ser Phe Glu Lys Leu Tyr Glu His Phe
65                  70                  75                  80

Leu Ser Gly Asp Lys Lys Ala Tyr Ala Lys Glu Ser Glu Arg Leu Lys
                85                  90                  95

Lys Glu Ile Val Lys Thr Leu Ile Lys Asn Leu Pro Glu Gly Ile Gly
            100                 105                 110

Lys Ile Ser Glu Ile Asn Ser Ala Lys Tyr Leu Asn Gly Val Leu Tyr
        115                 120                 125

Asp Phe Ile Asp Lys Thr His Lys Asp Ser Glu Glu Lys Gln Asn Ile
    130                 135                 140

Leu Ser Asp Ile Leu Glu Thr Lys Gly Tyr Leu Ala Leu Phe Ser Lys
145                 150                 155                 160

Phe Leu Thr Ser Arg Ile Thr Thr Leu Glu Gln Ser Met Pro Lys Arg
                165                 170                 175

Val Ile Glu Asn Phe Glu Ile Tyr Ala Ala Asn Ile Pro Lys Met Gln
            180                 185                 190

Asp Ala Leu Glu Arg Gly Ala Val Ser Phe Ala Ile Glu Tyr Glu Ser
        195                 200                 205

Ile Cys Ser Val Asp Tyr Tyr Asn Gln Ile Leu Ser Gln Glu Asp Ile
    210                 215                 220

Asp Ser Tyr Asn Arg Leu Ile Ser Gly Ile Met Asp Glu Asp Gly Ala
225                 230                 235                 240

Lys Glu Lys Gly Ile Asn Gln Thr Ile Ser Glu Lys Asn Ile Lys Ile
                245                 250                 255

Lys Ser Glu His Leu Glu Glu Lys Pro Phe Arg Ile Leu Lys Gln Leu
            260                 265                 270

His Lys Gln Ile Leu Glu Glu Arg Glu Lys Ala Phe Thr Ile Asp His
        275                 280                 285

Ile Asp Ser Asp Glu Glu Val Val Gln Val Thr Lys Glu Ala Phe Glu
    290                 295                 300
```

```
Gln Thr Lys Glu Gln Trp Glu Asn Ile Lys Lys Ile Asn Gly Phe Tyr
305                 310                 315                 320

Ala Lys Asp Pro Gly Asp Ile Thr Leu Phe Ile Val Val Gly Pro Asn
                325                 330                 335

Gln Thr His Val Leu Ser Gln Leu Ile Tyr Gly Glu His Asp Arg Ile
            340                 345                 350

Arg Leu Leu Leu Glu Glu Tyr Glu Lys Asn Thr Leu Glu Val Leu Pro
        355                 360                 365

Arg Arg Thr Lys Ser Glu Asp Ala Arg Tyr Asp Lys Phe Val Asn Ala
    370                 375                 380

Val Pro Lys Lys Val Ala Lys Glu Ser His Thr Phe Asp Gly Leu Gln
385                 390                 395                 400

Lys Met Thr Gly Asp Asp Arg Leu Phe Ile Leu Tyr Arg Asp Glu Leu
                405                 410                 415

Ala Arg Asn Tyr Met Arg Ile Lys Glu Ala Tyr Gly Thr Phe Glu Arg
            420                 425                 430

Asp Ile Leu Lys Ser Arg Arg Gly Ile Lys Gly Asn Arg Asp Val Gln
        435                 440                 445

Glu Ser Leu Val Ser Phe Tyr Asp Glu Leu Thr Lys Phe Arg Ser Ala
    450                 455                 460

Leu Arg Ile Ile Asn Ser Gly Asn Asp Glu Lys Ala Asp Pro Ile Phe
465                 470                 475                 480

Tyr Asn Thr Phe Asp Gly Ile Phe Glu Lys Ala Asn Arg Thr Tyr Lys
                485                 490                 495

Ala Glu Asn Leu Cys Arg Asn Tyr Val Thr Lys Ser Pro Ala Asp Asp
            500                 505                 510

Ala Arg Ile Met Ala Ser Cys Leu Gly Thr Pro Ala Arg Leu Arg Thr
        515                 520                 525

His Trp Trp Asn Gly Glu Glu Asn Phe Ala Ile Asn Asp Val Ala Met
    530                 535                 540

Ile Arg Arg Gly Asp Glu Tyr Tyr Tyr Phe Val Leu Thr Pro Asp Val
545                 550                 555                 560

Lys Pro Val Asp Leu Lys Thr Lys Asp Glu Thr Asp Ala Gln Ile Phe
                565                 570                 575

Val Gln Arg Lys Gly Ala Lys Ser Phe Leu Gly Leu Pro Lys Ala Leu
```

```
            580                 585                 590

Phe Lys Cys Ile Leu Glu Pro Tyr Phe Glu Ser Pro Glu His Lys Asn
        595                 600                 605

Asp Lys Asn Cys Val Ile Glu Glu Tyr Val Ser Lys Pro Leu Thr Ile
    610                 615                 620

Asp Arg Arg Ala Tyr Asp Ile Phe Lys Asn Gly Thr Phe Lys Lys Thr
625                 630                 635                 640

Asn Ile Gly Ile Asp Gly Leu Thr Glu Glu Lys Phe Lys Asp Asp Cys
                645                 650                 655

Arg Tyr Leu Ile Asp Val Tyr Lys Glu Phe Ile Ala Val Tyr Thr Arg
            660                 665                 670

Tyr Ser Cys Phe Asn Met Ser Gly Leu Lys Arg Ala Asp Glu Tyr Asn
        675                 680                 685

Asp Ile Gly Glu Phe Phe Ser Asp Val Asp Thr Arg Leu Cys Thr Met
    690                 695                 700

Glu Trp Ile Pro Val Ser Phe Glu Arg Ile Asn Asp Met Val Asp Lys
705                 710                 715                 720

Lys Glu Gly Leu Leu Phe Leu Val Arg Ser Met Phe Leu Tyr Asn Arg
                725                 730                 735

Pro Arg Lys Pro Tyr Glu Arg Thr Phe Ile Gln Leu Phe Ser Asp Ser
            740                 745                 750

Asn Met Glu His Thr Ser Met Leu Leu Asn Ser Arg Ala Met Ile Gln
        755                 760                 765

Tyr Arg Ala Ala Ser Leu Pro Arg Arg Val Thr His Lys Lys Gly Ser
    770                 775                 780

Ile Leu Val Ala Leu Arg Asp Ser Asn Gly Glu His Ile Pro Met His
785                 790                 795                 800

Ile Arg Glu Ala Ile Tyr Lys Met Lys Asn Asn Phe Asp Ile Ser Ser
                805                 810                 815

Glu Asp Phe Ile Met Ala Lys Ala Tyr Leu Ala Glu His Asp Val Ala
            820                 825                 830

Ile Lys Lys Ala Asn Glu Asp Ile Ile Arg Asn Arg Arg Tyr Thr Glu
        835                 840                 845

Asp Lys Phe Phe Leu Ser Leu Ser Tyr Thr Lys Asn Ala Asp Ile Ser
    850                 855                 860
```

```
Ala Arg Thr Leu Asp Tyr Ile Asn Asp Lys Val Glu Glu Asp Thr Gln
865                 870                 875                 880


Asp Ser Arg Met Ala Val Ile Val Thr Arg Asn Leu Lys Asp Leu Thr
                885                 890                 895


Tyr Val Ala Val Val Asp Glu Lys Asn Asn Val Leu Glu Glu Lys Ser
            900                 905                 910


Leu Asn Glu Ile Asp Gly Val Asn Tyr Arg Glu Leu Leu Lys Glu Arg
        915                 920                 925


Thr Lys Ile Lys Tyr His Asp Lys Thr Arg Leu Trp Gln Tyr Asp Val
    930                 935                 940


Ser Ser Lys Gly Leu Lys Glu Ala Tyr Val Glu Leu Ala Val Thr Gln
945                 950                 955                 960


Ile Ser Lys Leu Ala Thr Lys Tyr Asn Ala Val Val Val Val Glu Ser
                965                 970                 975


Met Ser Ser Thr Phe Lys Asp Lys Phe Ser Phe Leu Asp Glu Gln Ile
            980                 985                 990


Phe Lys Ala Phe Glu Ala Arg Leu  Cys Ala Arg Met Ser  Asp Leu Ser
        995                 1000                1005


Phe Asn  Thr Ile Lys Glu Gly  Glu Ala Gly Ser Ile  Ser Asn Pro
    1010                1015                1020


Ile Gln  Val Ser Asn Asn Asn  Gly Asn Ser Tyr Gln  Asp Gly Val
    1025                1030                1035


Ile Tyr  Phe Leu Asn Asn Ala  Tyr Thr Arg Thr Leu  Cys Pro Asp
    1040                1045                1050


Thr Gly  Phe Val Asp Val Phe  Asp Lys Thr Arg Leu  Ile Thr Met
    1055                1060                1065


Gln Ser  Lys Arg Gln Phe Phe  Ala Lys Met Lys Asp  Ile Arg Ile
    1070                1075                1080


Asp Asp  Gly Glu Met Leu Phe  Thr Phe Asn Leu Glu  Glu Tyr Pro
    1085                1090                1095


Thr Lys  Arg Leu Leu Asp Arg  Lys Glu Trp Thr Val  Lys Ile Ala
    1100                1105                1110


Gly Asp  Gly Ser Tyr Phe Asp  Lys Asp Lys Gly Glu  Tyr Val Tyr
    1115                1120                1125
```

```
Val Asn  Asp Ile Val Arg Glu  Gln Ile Ile Pro Ala  Leu Leu Glu
    1130                1135                1140


Asp Lys  Ala Val Phe Asp Gly  Asn Met Ala Glu Lys  Phe Leu Asp
    1145                1150                1155


Lys Thr  Ala Ile Ser Gly Lys  Ser Val Glu Leu Ile  Tyr Lys Trp
    1160                1165                1170


Phe Ala  Asn Ala Leu Tyr Gly  Ile Ile Thr Lys Lys  Asp Gly Glu
    1175                1180                1185


Lys Ile  Tyr Arg Ser Pro Ile  Thr Gly Thr Glu Ile  Asp Val Ser
    1190                1195                1200


Lys Asn  Thr Thr Tyr Asn Phe  Gly Lys Lys Phe Met  Phe Lys Gln
    1205                1210                1215


Glu Tyr  Arg Gly Asp Gly Asp  Phe Leu Asp Ala Phe  Leu Asn Tyr
    1220                1225                1230


Met Gln  Ala Gln Asp Ile Ala  Val
    1235                1240


<210>  6
<211>  1238
<212>  PRT
<213>  Candidatus Methanoplasma termitum

<400>  6

Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr
1               5                   10                  15


Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
            20                  25                  30


Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
        35                  40                  45


Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
    50                  55                  60


His Leu Thr Asn Met Ser Leu Asp Trp Asn Ser Leu Lys Gln Ile Ser
65                  70                  75                  80


Glu Lys Tyr Tyr Lys Ser Arg Glu Glu Lys Asp Lys Lys Val Phe Leu
                85                  90                  95


Ser Glu Gln Lys Arg Met Arg Gln Glu Ile Val Ser Glu Phe Lys Lys
            100                 105                 110


Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys Lys Leu Phe Ser Glu Leu
```

```
        115                 120                 125

Leu Lys Glu Glu Ile Tyr Lys Lys Gly Asn His Gln Glu Ile Asp Ala
    130                 135                 140

Leu Lys Ser Phe Asp Lys Phe Ser Gly Tyr Phe Ile Gly Leu His Glu
145                 150                 155                 160

Asn Arg Lys Asn Met Tyr Ser Asp Gly Asp Glu Ile Thr Ala Ile Ser
                165                 170                 175

Asn Arg Ile Val Asn Glu Asn Phe Pro Lys Phe Leu Asp Asn Leu Gln
            180                 185                 190

Lys Tyr Gln Glu Ala Arg Lys Lys Tyr Pro Glu Trp Ile Ile Lys Ala
        195                 200                 205

Glu Ser Ala Leu Val Ala His Asn Ile Lys Met Asp Ile Val Phe Ser
    210                 215                 220

Leu Glu Tyr Phe Asn Lys Val Leu Asn Gln Glu Gly Ile Gln Arg Tyr
225                 230                 235                 240

Asn Leu Ala Leu Gly Gly Tyr Val Thr Lys Ser Gly Glu Lys Met Met
                245                 250                 255

Gly Leu Asn Asp Ala Leu Asn Leu Ala His Gln Ser Glu Lys Ser Ser
            260                 265                 270

Lys Gly Arg Ile His Met Thr Pro Leu Phe Lys Gln Ile Leu Ser Glu
        275                 280                 285

Lys Glu Ser Phe Ser Tyr Ile Pro Asp Val Phe Thr Glu Asp Ser Gln
    290                 295                 300

Leu Leu Pro Ser Ile Gly Gly Phe Phe Ala Gln Ile Glu Asn Asp Lys
305                 310                 315                 320

Asp Gly Asn Ile Phe Asp Arg Ala Leu Glu Leu Ile Ser Ser Tyr Ala
                325                 330                 335

Glu Tyr Asp Thr Glu Arg Ile Tyr Ile Arg Gln Ala Asp Ile Asn Arg
            340                 345                 350

Val Ser Asn Val Ile Phe Gly Glu Trp Gly Thr Leu Gly Gly Leu Met
        355                 360                 365

Arg Glu Tyr Lys Ala Asp Ser Ile Asn Asp Ile Asn Leu Glu Arg Thr
    370                 375                 380

Cys Lys Lys Val Asp Lys Trp Leu Asp Ser Lys Glu Phe Ala Leu Ser
385                 390                 395                 400
```

```
Asp Val Leu Glu Ala Ile Asp Arg Thr Gly Asn Asn Asp Ala Phe Asn
                405                 410                 415

Glu Tyr Ile Ser Lys Met Arg Thr Ala Arg Glu Lys Ile Asp Ala Ala
            420                 425                 430

Arg Lys Glu Met Lys Phe Ile Ser Glu Lys Ile Ser Gly Asp Glu Glu
        435                 440                 445

Ser Ile His Ile Ile Lys Thr Leu Leu Asp Ser Val Gln Gln Phe Leu
    450                 455                 460

His Phe Phe Asn Leu Phe Lys Ala Arg Gln Asp Ile Pro Leu Asp Gly
465                 470                 475                 480

Ala Phe Tyr Ala Glu Phe Asp Glu Val His Ser Lys Leu Phe Ala Ile
                485                 490                 495

Val Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Lys Asn Asn Leu
            500                 505                 510

Asn Thr Lys Lys Ile Lys Leu Asn Phe Lys Asn Pro Thr Leu Ala Asn
        515                 520                 525

Gly Trp Asp Gln Asn Lys Val Tyr Asp Tyr Ala Ser Leu Ile Phe Leu
    530                 535                 540

Arg Asp Gly Asn Tyr Tyr Leu Gly Ile Ile Asn Pro Lys Arg Lys Lys
545                 550                 555                 560

Asn Ile Lys Phe Glu Gln Gly Ser Gly Asn Gly Pro Phe Tyr Arg Lys
                565                 570                 575

Met Val Tyr Lys Gln Ile Pro Gly Pro Asn Lys Asn Leu Arg Pro Val
            580                 585                 590

Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu Tyr Lys Pro Ser Lys Glu
        595                 600                 605

Ile Ile Glu Gly Tyr Glu Ala Asp Lys His Ile Arg Gly Asp Lys Phe
    610                 615                 620

Asp Leu Asp Phe Cys His Lys Leu Ile Asp Phe Phe Lys Glu Ser Ile
625                 630                 635                 640

Glu Lys His Lys Asp Trp Ser Lys Phe Asn Phe Tyr Phe Ser Pro Thr
                645                 650                 655

Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr Leu Asp Val Glu Lys Gln
            660                 665                 670
```

```
Gly Tyr Arg Met His Phe Glu Asn Ile Ser Ala Glu Thr Ile Asp Glu
        675                 680                 685

Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe Gln Ile Tyr Asn Lys Asp
    690                 695                 700

Phe Val Lys Ala Ala Thr Gly Lys Lys Asp Met His Thr Ile Tyr Trp
705                 710                 715                 720

Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln Asp Val Val Val Lys Leu
                725                 730                 735

Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp Lys Ser Asp Ile Lys Glu
            740                 745                 750

Ile Val His Arg Glu Gly Glu Ile Leu Val Asn Arg Thr Tyr Asn Gly
        755                 760                 765

Arg Thr Pro Val Pro Asp Lys Ile His Lys Lys Leu Thr Asp Tyr His
    770                 775                 780

Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala Lys Glu Tyr Leu Asp Lys
785                 790                 795                 800

Val Arg Tyr Phe Lys Ala His Tyr Asp Ile Thr Lys Asp Arg Arg Tyr
                805                 810                 815

Leu Asn Asp Lys Ile Tyr Phe His Val Pro Leu Thr Leu Asn Phe Lys
            820                 825                 830

Ala Asn Gly Lys Lys Asn Leu Asn Lys Met Val Ile Glu Lys Phe Leu
        835                 840                 845

Ser Asp Glu Lys Ala His Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn
    850                 855                 860

Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser Gly Lys Ile Ile Asp Gln
865                 870                 875                 880

Gln Ser Leu Asn Val Ile Asp Gly Phe Asp Tyr Arg Glu Lys Leu Asn
                885                 890                 895

Gln Arg Glu Ile Glu Met Lys Asp Ala Arg Gln Ser Trp Asn Ala Ile
            900                 905                 910

Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr Leu Ser Lys Ala Val His
        915                 920                 925

Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn Ala Ile Val Val Met Glu
    930                 935                 940
```

```
Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln
945                 950                 955                 960


Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile Asp Lys Met Asn Tyr Leu
                965                 970                 975


Val Phe Lys Asp Ala Pro Asp Glu Ser Pro Gly Gly Val Leu Asn Ala
            980                 985                 990


Tyr Gln Leu Thr Asn Pro Leu Glu  Ser Phe Ala Lys Leu  Gly Lys Gln
        995                 1000                1005


Thr Gly  Ile Leu Phe Tyr Val  Pro Ala Ala Tyr Thr  Ser Lys Ile
    1010                1015                1020


Asp Pro  Thr Thr Gly Phe Val  Asn Leu Phe Asn Thr  Ser Ser Lys
    1025                1030                1035


Thr Asn  Ala Gln Glu Arg Lys  Glu Phe Leu Gln Lys  Phe Glu Ser
    1040                1045                1050


Ile Ser  Tyr Ser Ala Lys Asp  Gly Gly Ile Phe Ala  Phe Ala Phe
    1055                1060                1065


Asp Tyr  Arg Lys Phe Gly Thr  Ser Lys Thr Asp His  Lys Asn Val
    1070                1075                1080


Trp Thr  Ala Tyr Thr Asn Gly  Glu Arg Met Arg Tyr  Ile Lys Glu
    1085                1090                1095


Lys Lys  Arg Asn Glu Leu Phe  Asp Pro Ser Lys Glu  Ile Lys Glu
    1100                1105                1110


Ala Leu  Thr Ser Ser Gly Ile  Lys Tyr Asp Gly Gly  Gln Asn Ile
    1115                1120                1125


Leu Pro  Asp Ile Leu Arg Ser  Asn Asn Asn Gly Leu  Ile Tyr Thr
    1130                1135                1140


Met Tyr  Ser Ser Phe Ile Ala  Ala Ile Gln Met Arg  Val Tyr Asp
    1145                1150                1155


Gly Lys  Glu Asp Tyr Ile Ile  Ser Pro Ile Lys Asn  Ser Lys Gly
    1160                1165                1170


Glu Phe  Phe Arg Thr Asp Pro  Lys Arg Arg Glu Leu  Pro Ile Asp
    1175                1180                1185


Ala Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala Leu Arg  Gly Glu Leu
    1190                1195                1200


Thr Met  Arg Ala Ile Ala Glu  Lys Phe Asp Pro Asp  Ser Glu Lys
```

```
    1205                1210                1215

Met Ala  Lys Leu Glu Leu Lys  His Lys Asp Trp Phe  Glu Phe Met
    1220                1225                1230

Gln Thr  Arg Gly Asp
    1235
```

<210> 7
<211> 1281
<212> PRT
<213> Eubacterium eligens

<400> 7

```
Met Asn Gly Asn Arg Ser Ile Val Tyr Arg Glu Phe Val Gly Val Ile
1               5                   10                  15

Pro Val Ala Lys Thr Leu Arg Asn Glu Leu Arg Pro Val Gly His Thr
            20                  25                  30

Gln Glu His Ile Ile Gln Asn Gly Leu Ile Gln Glu Asp Glu Leu Arg
        35                  40                  45

Gln Glu Lys Ser Thr Glu Leu Lys Asn Ile Met Asp Asp Tyr Tyr Arg
    50                  55                  60

Glu Tyr Ile Asp Lys Ser Leu Ser Gly Val Thr Asp Leu Asp Phe Thr
65                  70                  75                  80

Leu Leu Phe Glu Leu Met Asn Leu Val Gln Ser Ser Pro Ser Lys Asp
                85                  90                  95

Asn Lys Lys Ala Leu Glu Lys Glu Gln Ser Lys Met Arg Glu Gln Ile
            100                 105                 110

Cys Thr His Leu Gln Ser Asp Ser Asn Tyr Lys Asn Ile Phe Asn Ala
        115                 120                 125

Lys Leu Leu Lys Glu Ile Leu Pro Asp Phe Ile Lys Asn Tyr Asn Gln
    130                 135                 140

Tyr Asp Val Lys Asp Lys Ala Gly Lys Leu Glu Thr Leu Ala Leu Phe
145                 150                 155                 160

Asn Gly Phe Ser Thr Tyr Phe Thr Asp Phe Phe Glu Lys Arg Lys Asn
                165                 170                 175

Val Phe Thr Lys Glu Ala Val Ser Thr Ser Ile Ala Tyr Arg Ile Val
            180                 185                 190

His Glu Asn Ser Leu Ile Phe Leu Ala Asn Met Thr Ser Tyr Lys Lys
        195                 200                 205
```

```
Ile Ser Glu Lys Ala Leu Asp Glu Ile Glu Val Ile Glu Lys Asn Asn
    210                 215                 220

Gln Asp Lys Met Gly Asp Trp Glu Leu Asn Gln Ile Phe Asn Pro Asp
225                 230                 235                 240

Phe Tyr Asn Met Val Leu Ile Gln Ser Gly Ile Asp Phe Tyr Asn Glu
                245                 250                 255

Ile Cys Gly Val Val Asn Ala His Met Asn Leu Tyr Cys Gln Gln Thr
            260                 265                 270

Lys Asn Asn Tyr Asn Leu Phe Lys Met Arg Lys Leu His Lys Gln Ile
        275                 280                 285

Leu Ala Tyr Thr Ser Thr Ser Phe Glu Val Pro Lys Met Phe Glu Asp
    290                 295                 300

Asp Met Ser Val Tyr Asn Ala Val Asn Ala Phe Ile Asp Glu Thr Glu
305                 310                 315                 320

Lys Gly Asn Ile Ile Gly Lys Leu Lys Asp Ile Val Asn Lys Tyr Asp
                325                 330                 335

Glu Leu Asp Glu Lys Arg Ile Tyr Ile Ser Lys Asp Phe Tyr Glu Thr
            340                 345                 350

Leu Ser Cys Phe Met Ser Gly Asn Trp Asn Leu Ile Thr Gly Cys Val
        355                 360                 365

Glu Asn Phe Tyr Asp Glu Asn Ile His Ala Lys Gly Lys Ser Lys Glu
    370                 375                 380

Glu Lys Val Lys Lys Ala Val Lys Glu Asp Lys Tyr Lys Ser Ile Asn
385                 390                 395                 400

Asp Val Asn Asp Leu Val Glu Lys Tyr Ile Asp Glu Lys Glu Arg Asn
                405                 410                 415

Glu Phe Lys Asn Ser Asn Ala Lys Gln Tyr Ile Arg Glu Ile Ser Asn
            420                 425                 430

Ile Ile Thr Asp Thr Glu Thr Ala His Leu Glu Tyr Asp Asp His Ile
        435                 440                 445

Ser Leu Ile Glu Ser Glu Glu Lys Ala Asp Glu Met Lys Lys Arg Leu
    450                 455                 460

Asp Met Tyr Met Asn Met Tyr His Trp Ala Lys Ala Phe Ile Val Asp
465                 470                 475                 480
```

```
Glu Val Leu Asp Arg Asp Glu Met Phe Tyr Ser Asp Ile Asp Asp Ile
                485                 490                 495

Tyr Asn Ile Leu Glu Asn Ile Val Pro Leu Tyr Asn Arg Val Arg Asn
            500                 505                 510

Tyr Val Thr Gln Lys Pro Tyr Asn Ser Lys Lys Ile Lys Leu Asn Phe
        515                 520                 525

Gln Ser Pro Thr Leu Ala Asn Gly Trp Ser Gln Ser Lys Glu Phe Asp
    530                 535                 540

Asn Asn Ala Ile Ile Leu Ile Arg Asp Asn Lys Tyr Tyr Leu Ala Ile
545                 550                 555                 560

Phe Asn Ala Lys Asn Lys Pro Asp Lys Lys Ile Ile Gln Gly Asn Ser
                565                 570                 575

Asp Lys Lys Asn Asp Asn Asp Tyr Lys Lys Met Val Tyr Asn Leu Leu
            580                 585                 590

Pro Gly Ala Asn Lys Met Leu Pro Lys Val Phe Leu Ser Lys Lys Gly
        595                 600                 605

Ile Glu Thr Phe Lys Pro Ser Asp Tyr Ile Ile Ser Gly Tyr Asn Ala
    610                 615                 620

His Lys His Ile Lys Thr Ser Glu Asn Phe Asp Ile Ser Phe Cys Arg
625                 630                 635                 640

Asp Leu Ile Asp Tyr Phe Lys Asn Ser Ile Glu Lys His Ala Glu Trp
                645                 650                 655

Arg Lys Tyr Glu Phe Lys Phe Ser Ala Thr Asp Ser Tyr Ser Asp Ile
            660                 665                 670

Ser Glu Phe Tyr Arg Glu Val Glu Met Gln Gly Tyr Arg Ile Asp Trp
        675                 680                 685

Thr Tyr Ile Ser Glu Ala Asp Ile Asn Lys Leu Asp Glu Glu Gly Lys
    690                 695                 700

Ile Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Glu Asn Ser Thr
705                 710                 715                 720

Gly Lys Glu Asn Leu His Thr Met Tyr Phe Lys Asn Ile Phe Ser Glu
                725                 730                 735

Glu Asn Leu Asp Lys Ile Ile Lys Leu Asn Gly Gln Ala Glu Leu Phe
            740                 745                 750

Tyr Arg Arg Ala Ser Val Lys Asn Pro Val Lys His Lys Lys Asp Ser
```

```
        755                 760                 765

Val Leu Val Asn Lys Thr Tyr Lys Asn Gln Leu Asp Asn Gly Asp Val
    770                 775                 780

Val Arg Ile Pro Ile Pro Asp Asp Ile Tyr Asn Glu Ile Tyr Lys Met
785                 790                 795                 800

Tyr Asn Gly Tyr Ile Lys Glu Ser Asp Leu Ser Glu Ala Ala Lys Glu
                805                 810                 815

Tyr Leu Asp Lys Val Glu Val Arg Thr Ala Gln Lys Asp Ile Val Lys
            820                 825                 830

Asp Tyr Arg Tyr Thr Val Asp Lys Tyr Phe Ile His Thr Pro Ile Thr
        835                 840                 845

Ile Asn Tyr Lys Val Thr Ala Arg Asn Asn Val Asn Asp Met Val Val
    850                 855                 860

Lys Tyr Ile Ala Gln Asn Asp Asp Ile His Val Ile Gly Ile Asp Arg
865                 870                 875                 880

Gly Glu Arg Asn Leu Ile Tyr Ile Ser Val Ile Asp Ser His Gly Asn
                885                 890                 895

Ile Val Lys Gln Lys Ser Tyr Asn Ile Leu Asn Asn Tyr Asp Tyr Lys
            900                 905                 910

Lys Lys Leu Val Glu Lys Glu Lys Thr Arg Glu Tyr Ala Arg Lys Asn
        915                 920                 925

Trp Lys Ser Ile Gly Asn Ile Lys Glu Leu Lys Glu Gly Tyr Ile Ser
    930                 935                 940

Gly Val Val His Glu Ile Ala Met Leu Ile Val Glu Tyr Asn Ala Ile
945                 950                 955                 960

Ile Ala Met Glu Asp Leu Asn Tyr Gly Phe Lys Arg Gly Arg Phe Lys
                965                 970                 975

Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Ser Met Leu Ile Asn Lys
            980                 985                 990

Leu Asn Tyr Phe Ala Ser Lys Glu  Lys Ser Val Asp Glu  Pro Gly Gly
        995                 1000                1005

Leu Leu  Lys Gly Tyr Gln Leu  Thr Tyr Val Pro Asp  Asn Ile Lys
    1010                1015                1020

Asn Leu  Gly Lys Gln Cys Gly  Val Ile Phe Tyr Val  Pro Ala Ala
    1025                1030                1035
```

```
Phe Thr  Ser Lys Ile Asp Pro  Ser Thr Gly Phe Ile  Ser Ala Phe
    1040                1045                1050

Asn Phe  Lys Ser Ile Ser Thr  Asn Ala Ser Arg Lys  Gln Phe Phe
    1055                1060                1065

Met Gln  Phe Asp Glu Ile Arg  Tyr Cys Ala Glu Lys  Asp Met Phe
    1070                1075                1080

Ser Phe  Gly Phe Asp Tyr Asn  Asn Phe Asp Thr Tyr  Asn Ile Thr
    1085                1090                1095

Met Gly  Lys Thr Gln Trp Thr  Val Tyr Thr Asn Gly  Glu Arg Leu
    1100                1105                1110

Gln Ser  Glu Phe Asn Asn Ala  Arg Arg Thr Gly Lys  Thr Lys Ser
    1115                1120                1125

Ile Asn  Leu Thr Glu Thr Ile  Lys Leu Leu Leu Glu  Asp Asn Glu
    1130                1135                1140

Ile Asn  Tyr Ala Asp Gly His  Asp Ile Arg Ile Asp  Met Glu Lys
    1145                1150                1155

Met Asp  Glu Asp Lys Lys Ser  Glu Phe Phe Ala Gln  Leu Leu Ser
    1160                1165                1170

Leu Tyr  Lys Leu Thr Val Gln  Met Arg Asn Ser Tyr  Thr Glu Ala
    1175                1180                1185

Glu Glu  Gln Glu Asn Gly Ile  Ser Tyr Asp Lys Ile  Ile Ser Pro
    1190                1195                1200

Val Ile  Asn Asp Glu Gly Glu  Phe Phe Asp Ser Asp  Asn Tyr Lys
    1205                1210                1215

Glu Ser  Asp Asp Lys Glu Cys  Lys Met Pro Lys Asp  Ala Asp Ala
    1220                1225                1230

Asn Gly  Ala Tyr Cys Ile Ala  Leu Lys Gly Leu Tyr  Glu Val Leu
    1235                1240                1245

Lys Ile  Lys Ser Glu Trp Thr  Glu Asp Gly Phe Asp  Arg Asn Cys
    1250                1255                1260

Leu Lys  Leu Pro His Ala Glu  Trp Leu Asp Phe Ile  Gln Asn Lys
    1265                1270                1275

Arg Tyr  Glu
    1280
```

<210> 8
<211> 1300
<212> PRT
<213> Francisella novicida

<400> 8

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1   5   10   15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
            20   25   30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
        35   40   45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
    50   55   60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65   70   75   80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85   90   95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
            100   105   110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
        115   120   125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
    130   135   140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145   150   155   160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
                165   170   175

Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Val Asn Tyr Ser Ser
            180   185   190

Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
        195   200   205

Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
    210   215   220

Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225   230   235   240

Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg

```
                245                 250                 255

Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
            260                 265                 270

Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
        275                 280                 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
    290                 295                 300

Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305                 310                 315                 320

Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
                325                 330                 335

Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
            340                 345                 350

Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
        355                 360                 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
    370                 375                 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385                 390                 395                 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
                405                 410                 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
            420                 425                 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
        435                 440                 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
    450                 455                 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465                 470                 475                 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
                485                 490                 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
            500                 505                 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
        515                 520                 525
```

```
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
    530                 535                 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545                 550                 555                 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
                565                 570                 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
            580                 585                 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
        595                 600                 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
    610                 615                 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625                 630                 635                 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
                645                 650                 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
            660                 665                 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
        675                 680                 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
    690                 695                 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705                 710                 715                 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
                725                 730                 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
            740                 745                 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
        755                 760                 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
    770                 775                 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785                 790                 795                 800
```

```
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
                805                 810                 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
            820                 825                 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
        835                 840                 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
    850                 855                 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865                 870                 875                 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
                885                 890                 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
            900                 905                 910

Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
        915                 920                 925

Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
    930                 935                 940

Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945                 950                 955                 960

Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
                965                 970                 975

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
            980                 985                 990

Ala Lys Leu Val Ile Glu Tyr Asn  Ala Ile Val Val Phe  Glu Asp Leu
        995                 1000                1005

Asn Phe  Gly Phe Lys Arg Gly  Arg Phe Lys Val Glu  Lys Gln Val
    1010                1015                1020

Tyr Gln  Lys Leu Glu Lys Met  Leu Ile Glu Lys Leu  Asn Tyr Leu
    1025                1030                1035

Val Phe  Lys Asp Asn Glu Phe  Asp Lys Thr Gly Gly  Val Leu Arg
    1040                1045                1050

Ala Tyr  Gln Leu Thr Ala Pro  Phe Glu Thr Phe Lys  Lys Met Gly
    1055                1060                1065
```

```
Lys Gln  Thr Gly Ile Ile Tyr  Tyr Val Pro Ala Gly  Phe Thr Ser
    1070                1075                1080

Lys Ile  Cys Pro Val Thr Gly  Phe Val Asn Gln Leu  Tyr Pro Lys
    1085                1090                1095

Tyr Glu  Ser Val Ser Lys Ser  Gln Glu Phe Phe Ser  Lys Phe Asp
    1100                1105                1110

Lys Ile  Cys Tyr Asn Leu Asp  Lys Gly Tyr Phe Glu  Phe Ser Phe
    1115                1120                1125

Asp Tyr  Lys Asn Phe Gly Asp  Lys Ala Ala Lys Gly  Lys Trp Thr
    1130                1135                1140

Ile Ala  Ser Phe Gly Ser Arg  Leu Ile Asn Phe Arg  Asn Ser Asp
    1145                1150                1155

Lys Asn  His Asn Trp Asp Thr  Arg Glu Val Tyr Pro  Thr Lys Glu
    1160                1165                1170

Leu Glu  Lys Leu Leu Lys Asp  Tyr Ser Ile Glu Tyr  Gly His Gly
    1175                1180                1185

Glu Cys  Ile Lys Ala Ala Ile  Cys Gly Glu Ser Asp  Lys Lys Phe
    1190                1195                1200

Phe Ala  Lys Leu Thr Ser Val  Leu Asn Thr Ile Leu  Gln Met Arg
    1205                1210                1215

Asn Ser  Lys Thr Gly Thr Glu  Leu Asp Tyr Leu Ile  Ser Pro Val
    1220                1225                1230

Ala Asp  Val Asn Gly Asn Phe  Phe Asp Ser Arg Gln  Ala Pro Lys
    1235                1240                1245

Asn Met  Pro Gln Asp Ala Asp  Ala Asn Gly Ala Tyr  His Ile Gly
    1250                1255                1260

Leu Lys  Gly Leu Met Leu Leu  Gly Arg Ile Lys Asn  Asn Gln Glu
    1265                1270                1275

Gly Lys  Lys Leu Asn Leu Val  Ile Lys Asn Glu Glu  Tyr Phe Glu
    1280                1285                1290

Phe Val  Gln Asn Arg Asn Asn
    1295                1300
```

<210> 9
<211> 1206
<212> PRT
<213> Unknown

<220>
<223> Lachnospiraceae bacterium

<400> 9

Met Tyr Tyr Glu Ser Leu Thr Lys Gln Tyr Pro Val Ser Lys Thr Ile
1   5   10   15

Arg Asn Glu Leu Ile Pro Ile Gly Lys Thr Leu Asp Asn Ile Arg Gln
            20   25   30

Asn Asn Ile Leu Glu Ser Asp Val Lys Arg Lys Gln Asn Tyr Glu His
        35   40   45

Val Lys Gly Ile Leu Asp Glu Tyr His Lys Gln Leu Ile Asn Glu Ala
    50   55   60

Leu Asp Asn Cys Thr Leu Pro Ser Leu Lys Ile Ala Ala Glu Ile Tyr
65   70   75   80

Leu Lys Asn Gln Lys Glu Val Ser Asp Arg Glu Asp Phe Asn Lys Thr
                85   90   95

Gln Asp Leu Leu Arg Lys Glu Val Val Glu Lys Leu Lys Ala His Glu
            100   105   110

Asn Phe Thr Lys Ile Gly Lys Lys Asp Ile Leu Asp Leu Leu Glu Lys
        115   120   125

Leu Pro Ser Ile Ser Glu Asp Asp Tyr Asn Ala Leu Glu Ser Phe Arg
    130   135   140

Asn Phe Tyr Thr Tyr Phe Thr Ser Tyr Asn Lys Val Arg Glu Asn Leu
145   150   155   160

Tyr Ser Asp Lys Glu Lys Ser Ser Thr Val Ala Tyr Arg Leu Ile Asn
                165   170   175

Glu Asn Phe Pro Lys Phe Leu Asp Asn Val Lys Ser Tyr Arg Phe Val
            180   185   190

Lys Thr Ala Gly Ile Leu Ala Asp Gly Leu Gly Glu Glu Glu Gln Asp
        195   200   205

Ser Leu Phe Ile Val Glu Thr Phe Asn Lys Thr Leu Thr Gln Asp Gly
    210   215   220

Ile Asp Thr Tyr Asn Ser Gln Val Gly Lys Ile Asn Ser Ser Ile Asn
225   230   235   240

Leu Tyr Asn Gln Lys Asn Gln Lys Ala Asn Gly Phe Arg Lys Ile Pro
                245   250   255

```
Lys Met Lys Met Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Glu Ser
            260                 265                 270

Phe Ile Asp Glu Phe Gln Ser Asp Glu Val Leu Ile Asp Asn Val Glu
        275                 280                 285

Ser Tyr Gly Ser Val Leu Ile Glu Ser Leu Lys Ser Ser Lys Val Ser
    290                 295                 300

Ala Phe Phe Asp Ala Leu Arg Glu Ser Lys Gly Lys Asn Val Tyr Val
305                 310                 315                 320

Lys Asn Asp Leu Ala Lys Thr Ala Met Ser Val Ile Val Phe Glu Asn
                325                 330                 335

Trp Arg Thr Phe Asp Asp Leu Leu Asn Gln Glu Tyr Asp Leu Ala Asn
            340                 345                 350

Glu Asn Lys Lys Lys Asp Asp Lys Tyr Phe Glu Lys Arg Gln Lys Glu
        355                 360                 365

Leu Lys Lys Asn Lys Ser Tyr Ser Leu Glu His Leu Cys Asn Leu Ser
    370                 375                 380

Glu Asp Ser Cys Asn Leu Ile Glu Asn Tyr Ile His Gln Ile Ser Asp
385                 390                 395                 400

Asp Ile Glu Asn Ile Ile Ile Asn Asn Glu Thr Phe Leu Arg Ile Val
                405                 410                 415

Ile Asn Glu His Asp Arg Ser Arg Lys Leu Ala Lys Asn Arg Lys Ala
            420                 425                 430

Val Lys Ala Ile Lys Asp Phe Leu Asp Ser Ile Lys Val Leu Glu Arg
        435                 440                 445

Glu Leu Lys Leu Ile Asn Ser Ser Gly Gln Glu Leu Glu Lys Asp Leu
    450                 455                 460

Ile Val Tyr Ser Ala His Glu Glu Leu Leu Val Glu Leu Lys Gln Val
465                 470                 475                 480

Asp Ser Leu Tyr Asn Met Thr Arg Asn Tyr Leu Thr Lys Lys Pro Phe
                485                 490                 495

Ser Thr Glu Lys Val Lys Leu Asn Phe Asn Arg Ser Thr Leu Leu Asn
            500                 505                 510

Gly Trp Asp Arg Asn Lys Glu Thr Asp Asn Leu Gly Val Leu Leu Leu
        515                 520                 525

Lys Asp Gly Lys Tyr Tyr Leu Gly Ile Met Asn Thr Ser Ala Asn Lys
```

```
    530                 535                 540

Ala Phe Val Asn Pro Pro Val Ala Lys Thr Glu Lys Val Phe Lys Lys
545                 550                 555                 560

Val Asp Tyr Lys Leu Leu Pro Val Pro Asn Gln Met Leu Pro Lys Val
                565                 570                 575

Phe Phe Ala Lys Ser Asn Ile Asp Phe Tyr Asn Pro Ser Ser Glu Ile
            580                 585                 590

Tyr Ser Asn Tyr Lys Lys Gly Thr His Lys Lys Gly Asn Met Phe Ser
        595                 600                 605

Leu Glu Asp Cys His Asn Leu Ile Asp Phe Phe Lys Glu Ser Ile Ser
    610                 615                 620

Lys His Glu Asp Trp Ser Lys Phe Gly Phe Lys Phe Asp Thr Gln Ala
625                 630                 635                 640

Ser Tyr Asn Asp Ile Ser Glu Phe Tyr Arg Glu Val Glu Lys Gln Gly
                645                 650                 655

Tyr Lys Leu Thr Tyr Thr Asp Ile Asp Glu Thr Tyr Ile Asn Asp Leu
            660                 665                 670

Ile Glu Arg Asn Glu Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
        675                 680                 685

Ser Met Tyr Ser Lys Gly Lys Leu Asn Leu His Thr Leu Tyr Phe Met
    690                 695                 700

Met Leu Phe Asp Gln Arg Asn Ile Asp Asp Val Val Tyr Lys Leu Asn
705                 710                 715                 720

Gly Glu Ala Glu Val Phe Tyr Arg Pro Ala Ser Ile Ser Glu Asp Glu
                725                 730                 735

Leu Ile Ile His Lys Ala Gly Glu Glu Ile Lys Asn Lys Asn Pro Asn
            740                 745                 750

Arg Ala Arg Thr Lys Glu Thr Ser Thr Phe Ser Tyr Asp Ile Val Lys
        755                 760                 765

Asp Lys Arg Tyr Ser Lys Asp Lys Phe Thr Leu His Ile Pro Ile Thr
    770                 775                 780

Met Asn Phe Gly Val Asp Glu Val Lys Arg Phe Asn Asp Ala Val Asn
785                 790                 795                 800

Ser Ala Ile Arg Ile Asp Glu Asn Val Asn Val Ile Gly Ile Asp Arg
                805                 810                 815
```

```
Gly Glu Arg Asn Leu Leu Tyr Val Val Val Ile Asp Ser Lys Gly Asn
            820                 825                 830

Ile Leu Glu Gln Ile Ser Leu Asn Ser Ile Ile Asn Lys Glu Tyr Asp
        835                 840                 845

Ile Glu Thr Asp Tyr His Ala Leu Leu Asp Glu Arg Glu Gly Gly Arg
    850                 855                 860

Asp Lys Ala Arg Lys Asp Trp Asn Thr Val Glu Asn Ile Arg Asp Leu
865                 870                 875                 880

Lys Ala Gly Leu Tyr Leu Gln Val Val Asn Val Val Ala Lys Leu Val
                885                 890                 895

Leu Lys Tyr Asn Ala Ile Ile Cys Leu Glu Asp Leu Asn Phe Gly Phe
            900                 905                 910

Lys Arg Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu
        915                 920                 925

Lys Met Leu Ile Asp Lys Leu Asn Tyr Leu Val Ile Asp Lys Ser Arg
    930                 935                 940

Glu Gln Thr Ser Pro Lys Glu Leu Gly Gly Ala Leu Asn Ala Leu Gln
945                 950                 955                 960

Leu Thr Ser Lys Phe Lys Ser Phe Lys Glu Leu Gly Lys Gln Ser Gly
                965                 970                 975

Val Ile Tyr Tyr Val Pro Ala Tyr Leu Thr Ser Lys Ile Asp Pro Thr
            980                 985                 990

Thr Gly Phe Ala Asn Leu Phe Tyr  Met Lys Cys Glu Asn  Val Glu Lys
        995                 1000                1005

Ser Lys  Arg Phe Phe Asp Gly  Phe Asp Phe Ile Arg  Phe Asn Ala
    1010                1015                1020

Leu Glu  Asn Val Phe Glu Phe  Gly Phe Asp Tyr Arg  Ser Phe Thr
    1025                1030                1035

Gln Arg  Ala Cys Gly Ile Asn  Ser Lys Trp Thr Val  Cys Thr Asn
    1040                1045                1050

Gly Glu  Arg Ile Ile Lys Tyr  Arg Asn Pro Asp Lys  Asn Asn Met
    1055                1060                1065

Phe Asp  Glu Lys Val Val Val  Val Thr Asp Glu Met  Lys Asn Leu
    1070                1075                1080
```

```
Phe Glu  Gln Tyr Lys Ile Pro  Tyr Glu Asp Gly Arg  Asn Val Lys
    1085                1090                1095


Asp Met  Ile Ile Ser Asn Glu  Glu Ala Glu Phe Tyr  Arg Arg Leu
    1100                1105                1110


Tyr Arg  Leu Leu Gln Gln Thr  Leu Gln Met Arg Asn  Ser Thr Ser
    1115                1120                1125


Asp Gly  Thr Arg Asp Tyr Ile  Ile Ser Pro Val Lys  Asn Lys Arg
    1130                1135                1140


Glu Ala  Tyr Phe Asn Ser Glu  Leu Ser Asp Gly Ser  Val Pro Lys
    1145                1150                1155


Asp Ala  Asp Ala Asn Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Leu
    1160                1165                1170


Trp Val  Leu Glu Gln Ile Arg  Gln Lys Ser Glu Gly  Glu Lys Ile
    1175                1180                1185


Asn Leu  Ala Met Thr Asn Ala  Glu Trp Leu Glu Tyr  Ala Gln Thr
    1190                1195                1200


His Leu  Leu
    1205


<210>  10
<211>  1233
<212>  PRT
<213>  Unknown

<220>
<223>  Lachnospiraceae bacterium

<400>  10

Met Asp Tyr Gly Asn Gly Gln Phe Glu Arg Arg Ala Pro Leu Thr Lys
1               5                   10                  15


Thr Ile Thr Leu Arg Leu Lys Pro Ile Gly Glu Thr Arg Glu Thr Ile
            20                  25                  30


Arg Glu Gln Lys Leu Leu Glu Gln Asp Ala Ala Phe Arg Lys Leu Val
        35                  40                  45


Glu Thr Val Thr Pro Ile Val Asp Asp Cys Ile Arg Lys Ile Ala Asp
    50                  55                  60


Asn Ala Leu Cys His Phe Gly Thr Glu Tyr Asp Phe Ser Cys Leu Gly
65                  70                  75                  80


Asn Ala Ile Ser Lys Asn Asp Ser Lys Ala Ile Lys Lys Glu Thr Glu
                85                  90                  95
```

```
Lys Val Glu Lys Leu Leu Ala Lys Val Leu Thr Glu Asn Leu Pro Asp
            100                 105                 110

Gly Leu Arg Lys Val Asn Asp Ile Asn Ser Ala Ala Phe Ile Gln Asp
        115                 120                 125

Thr Leu Thr Ser Phe Val Gln Asp Asp Ala Asp Lys Arg Val Leu Ile
    130                 135                 140

Gln Glu Leu Lys Gly Lys Thr Val Leu Met Gln Arg Phe Leu Thr Thr
145                 150                 155                 160

Arg Ile Thr Ala Leu Thr Val Trp Leu Pro Asp Arg Val Phe Glu Asn
                165                 170                 175

Phe Asn Ile Phe Ile Glu Asn Ala Glu Lys Met Arg Ile Leu Leu Asp
            180                 185                 190

Ser Pro Leu Asn Glu Lys Ile Met Lys Phe Asp Pro Asp Ala Glu Gln
        195                 200                 205

Tyr Ala Ser Leu Glu Phe Tyr Gly Gln Cys Leu Ser Gln Lys Asp Ile
    210                 215                 220

Asp Ser Tyr Asn Leu Ile Ile Ser Gly Ile Tyr Ala Asp Asp Glu Val
225                 230                 235                 240

Lys Asn Pro Gly Ile Asn Glu Ile Val Lys Glu Tyr Asn Gln Gln Ile
                245                 250                 255

Arg Gly Asp Lys Asp Glu Ser Pro Leu Pro Lys Leu Lys Lys Leu His
            260                 265                 270

Lys Gln Ile Leu Met Pro Val Glu Lys Ala Phe Phe Val Arg Val Leu
        275                 280                 285

Ser Asn Asp Ser Asp Ala Arg Ser Ile Leu Glu Lys Ile Leu Lys Asp
    290                 295                 300

Thr Glu Met Leu Pro Ser Lys Ile Ile Glu Ala Met Lys Glu Ala Asp
305                 310                 315                 320

Ala Gly Asp Ile Ala Val Tyr Gly Ser Arg Leu His Glu Leu Ser His
                325                 330                 335

Val Ile Tyr Gly Asp His Gly Lys Leu Ser Gln Ile Ile Tyr Asp Lys
            340                 345                 350

Glu Ser Lys Arg Ile Ser Glu Leu Met Glu Thr Leu Ser Pro Lys Glu
        355                 360                 365
```

```
Arg Lys Glu Ser Lys Lys Arg Leu Glu Gly Leu Glu Glu His Ile Arg
    370                 375                 380

Lys Ser Thr Tyr Thr Phe Asp Glu Leu Asn Arg Tyr Ala Glu Lys Asn
385                 390                 395                 400

Val Met Ala Ala Tyr Ile Ala Ala Val Glu Glu Ser Cys Ala Glu Ile
                405                 410                 415

Met Arg Lys Glu Lys Asp Leu Arg Thr Leu Leu Ser Lys Glu Asp Val
            420                 425                 430

Lys Ile Arg Gly Asn Arg His Asn Thr Leu Ile Val Lys Asn Tyr Phe
        435                 440                 445

Asn Ala Trp Thr Val Phe Arg Asn Leu Ile Arg Ile Leu Arg Arg Lys
    450                 455                 460

Ser Glu Ala Glu Ile Asp Ser Asp Phe Tyr Asp Val Leu Asp Asp Ser
465                 470                 475                 480

Val Glu Val Leu Ser Leu Thr Tyr Lys Gly Glu Asn Leu Cys Arg Ser
                485                 490                 495

Tyr Ile Thr Lys Lys Ile Gly Ser Asp Leu Lys Pro Glu Ile Ala Thr
            500                 505                 510

Tyr Gly Ser Ala Leu Arg Pro Asn Ser Arg Trp Trp Ser Pro Gly Glu
        515                 520                 525

Lys Phe Asn Val Lys Phe His Thr Ile Val Arg Arg Asp Gly Arg Leu
    530                 535                 540

Tyr Tyr Phe Ile Leu Pro Lys Gly Ala Lys Pro Val Glu Leu Glu Asp
545                 550                 555                 560

Met Asp Gly Asp Ile Glu Cys Leu Gln Met Arg Lys Ile Pro Asn Pro
                565                 570                 575

Thr Ile Phe Leu Pro Lys Leu Val Phe Lys Asp Pro Glu Ala Phe Phe
            580                 585                 590

Arg Asp Asn Pro Glu Ala Asp Glu Phe Val Phe Leu Ser Gly Met Lys
        595                 600                 605

Ala Pro Val Thr Ile Thr Arg Glu Thr Tyr Glu Ala Tyr Arg Tyr Lys
    610                 615                 620

Leu Tyr Thr Val Gly Lys Leu Arg Asp Gly Glu Val Ser Glu Glu Glu
625                 630                 635                 640
```

```
Tyr Lys Arg Ala Leu Leu Gln Val Leu Thr Ala Tyr Lys Glu Phe Leu
                645                 650                 655

Glu Asn Arg Met Ile Tyr Ala Asp Leu Asn Phe Gly Phe Lys Asp Leu
            660                 665                 670

Glu Glu Tyr Lys Asp Ser Ser Glu Phe Ile Lys Gln Val Glu Thr His
        675                 680                 685

Asn Thr Phe Met Cys Trp Ala Lys Val Ser Ser Ser Gln Leu Asp Asp
    690                 695                 700

Leu Val Lys Ser Gly Asn Gly Leu Leu Phe Glu Ile Trp Ser Glu Arg
705                 710                 715                 720

Leu Glu Ser Tyr Tyr Lys Tyr Gly Asn Glu Lys Val Leu Arg Gly Tyr
                725                 730                 735

Glu Gly Val Leu Leu Ser Ile Leu Lys Asp Glu Asn Leu Val Ser Met
            740                 745                 750

Arg Thr Leu Leu Asn Ser Arg Pro Met Leu Val Tyr Arg Pro Lys Glu
        755                 760                 765

Ser Ser Lys Pro Met Val Val His Arg Asp Gly Ser Arg Val Val Asp
    770                 775                 780

Arg Phe Asp Lys Asp Gly Lys Tyr Ile Pro Pro Glu Val His Asp Glu
785                 790                 795                 800

Leu Tyr Arg Phe Phe Asn Asn Leu Leu Ile Lys Glu Lys Leu Gly Glu
                805                 810                 815

Lys Ala Arg Lys Ile Leu Asp Asn Lys Lys Val Lys Val Lys Val Leu
            820                 825                 830

Glu Ser Glu Arg Val Lys Trp Ser Lys Phe Tyr Asp Glu Gln Phe Ala
        835                 840                 845

Val Thr Phe Ser Val Lys Lys Asn Ala Asp Cys Leu Asp Thr Thr Lys
    850                 855                 860

Asp Leu Asn Ala Glu Val Met Glu Gln Tyr Ser Glu Ser Asn Arg Leu
865                 870                 875                 880

Ile Leu Ile Arg Asn Thr Thr Asp Ile Leu Tyr Tyr Leu Val Leu Asp
                885                 890                 895

Lys Asn Gly Lys Val Leu Lys Gln Arg Ser Leu Asn Ile Ile Asn Asp
            900                 905                 910

Gly Ala Arg Asp Val Asp Trp Lys Glu Arg Phe Arg Gln Val Thr Lys
```

```
        915                 920                 925

Asp Arg Asn Glu Gly Tyr Asn Glu Trp Asp Tyr Ser Arg Thr Ser Asn
    930                 935                 940

Asp Leu Lys Glu Val Tyr Leu Asn Tyr Ala Leu Lys Glu Ile Ala Glu
945                 950                 955                 960

Ala Val Ile Glu Tyr Asn Ala Ile Leu Ile Ile Glu Lys Met Ser Asn
                965                 970                 975

Ala Phe Lys Asp Lys Tyr Ser Phe Leu Asp Asp Val Thr Phe Lys Gly
            980                 985                 990

Phe Glu Thr Lys Lys Leu Ala Lys  Leu Ser Asp Leu His  Phe Arg Gly
        995                 1000                1005

Ile Lys  Asp Gly Glu Pro Cys  Ser Phe Thr Asn Pro  Leu Gln Leu
    1010                1015                1020

Cys Gln  Asn Asp Ser Asn Lys  Ile Leu Gln Asp Gly  Val Ile Phe
    1025                1030                1035

Met Val  Pro Asn Ser Met Thr  Arg Ser Leu Asp Pro  Asp Thr Gly
    1040                1045                1050

Phe Ile  Phe Ala Ile Asn Asp  His Asn Ile Arg Thr  Lys Lys Ala
    1055                1060                1065

Lys Leu  Asn Phe Leu Ser Lys  Phe Asp Gln Leu Lys  Val Ser Ser
    1070                1075                1080

Glu Gly  Cys Leu Ile Met Lys  Tyr Ser Gly Asp Ser  Leu Pro Thr
    1085                1090                1095

His Asn  Thr Asp Asn Arg Val  Trp Asn Cys Cys Cys  Asn His Pro
    1100                1105                1110

Ile Thr  Asn Tyr Asp Arg Glu  Thr Lys Lys Val Glu  Phe Ile Glu
    1115                1120                1125

Glu Pro  Val Glu Glu Leu Ser  Arg Val Leu Glu Glu  Asn Gly Ile
    1130                1135                1140

Glu Thr  Asp Thr Glu Leu Asn  Lys Leu Asn Glu Arg  Glu Asn Val
    1145                1150                1155

Pro Gly  Lys Val Val Asp Ala  Ile Tyr Ser Leu Val  Leu Asn Tyr
    1160                1165                1170

Leu Arg  Gly Thr Val Ser Gly  Val Ala Gly Gln Arg  Ala Val Tyr
    1175                1180                1185
```

```
Tyr Ser  Pro Val Thr Gly Lys  Lys Tyr Asp Ile Ser  Phe Ile Gln
    1190                1195                1200


Ala Met  Asn Leu Asn Arg Lys  Cys Asp Tyr Tyr Arg  Ile Gly Ser
    1205                1210                1215


Lys Glu  Arg Gly Glu Trp Thr  Asp Phe Val Ala Gln  Leu Ile Asn
    1220                1225                1230


<210>  11
<211>  1227
<212>  PRT
<213>  Unknown

<220>
<223>  Lachnospiraceae bacterium

<400>  11

Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1               5                   10                  15


Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
            20                  25                  30


Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
        35                  40                  45


Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
    50                  55                  60


Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile Ser Leu
65                  70                  75                  80


Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu Glu Asn
                85                  90                  95


Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys Gly Asn
            100                 105                 110


Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr Ile Leu
        115                 120                 125


Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn Ser Phe
    130                 135                 140


Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg Glu Asn
145                 150                 155                 160


Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg Cys Ile
                165                 170                 175


Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe Glu Lys
```

```
            180                 185                 190

Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys Glu Lys
        195                 200                 205

Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly Glu Phe
    210                 215                 220

Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn Ala Ile
225                 230                 235                 240

Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly Leu Asn
                245                 250                 255

Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu Pro Lys
            260                 265                 270

Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser Leu Ser
        275                 280                 285

Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu Val Phe
    290                 295                 300

Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile Lys Lys
305                 310                 315                 320

Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala Gly Ile
                325                 330                 335

Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp Ile Phe
            340                 345                 350

Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr Asp Asp
        355                 360                 365

Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu Asp Asp
    370                 375                 380

Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu Gln Leu
385                 390                 395                 400

Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu Lys Glu
                405                 410                 415

Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly Ser Ser
            420                 425                 430

Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu Lys Lys
        435                 440                 445

Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser Val Lys
    450                 455                 460
```

```
Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys Glu Thr
465                 470                 475                 480

Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr Asp Ile
                485                 490                 495

Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr Val Thr
            500                 505                 510

Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln Asn Pro
        515                 520                 525

Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr Arg Ala
    530                 535                 540

Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met Asp Lys
545                 550                 555                 560

Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val Asn Gly
                565                 570                 575

Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn Lys Met
            580                 585                 590

Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr Asn Pro
        595                 600                 605

Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys Lys Gly
    610                 615                 620

Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe Phe Lys
625                 630                 635                 640

Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp Phe Asn
                645                 650                 655

Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr Arg Glu
            660                 665                 670

Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser Lys Lys
        675                 680                 685

Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe Gln Ile
    690                 695                 700

Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn Leu His
705                 710                 715                 720

Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly Gln Ile
                725                 730                 735
```

```
Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser Leu Lys
            740                 745                 750

Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala Asn Lys
        755                 760                 765

Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp Val Tyr
    770                 775                 780

Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile Pro Ile
785                 790                 795                 800

Ala Asn Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr Glu
                805                 810                 815

Val Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly Ile
            820                 825                 830

Asp Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly Lys
        835                 840                 845

Gly Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn Phe
    850                 855                 860

Asn Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys Lys
865                 870                 875                 880

Glu Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu Asn
                885                 890                 895

Ile Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys Ile
            900                 905                 910

Cys Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp Leu
        915                 920                 925

Asn Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val Tyr
    930                 935                 940

Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val Asp
945                 950                 955                 960

Lys Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr Gln
                965                 970                 975

Ile Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn Gly
            980                 985                 990

Phe Ile Phe Tyr Ile Pro Ala Trp  Leu Thr Ser Lys Ile  Asp Pro Ser
        995                 1000                1005
```

```
Thr Gly  Phe Val Asn Leu Leu  Lys Thr Lys Tyr Thr  Ser Ile Ala
    1010                1015                1020


Asp Lys  Lys Phe Ile Ser Ser  Phe Asp Arg Ile Met  Tyr Val Pro
    1025                1030                1035


Glu Glu  Asp Leu Phe Glu Phe  Ala Leu Asp Tyr Lys  Asn Phe Ser
    1040                1045                1050


Arg Thr  Asp Ala Asp Tyr Ile  Lys Lys Trp Lys Leu  Tyr Ser Tyr
    1055                1060                1065


Gly Asn  Arg Ile Arg Ile Phe  Arg Asn Pro Lys Lys  Asn Asn Val
    1070                1075                1080


Phe Asp  Trp Glu Glu Val Cys  Leu Thr Ser Ala Tyr  Lys Glu Leu
    1085                1090                1095


Phe Asn  Lys Tyr Gly Ile Asn  Tyr Gln Gln Gly Asp  Ile Arg Ala
    1100                1105                1110


Leu Leu  Cys Glu Gln Ser Asp  Lys Ala Phe Tyr Ser  Ser Phe Met
    1115                1120                1125


Ala Leu  Met Ser Leu Met Leu  Gln Met Arg Asn Ser  Ile Thr Gly
    1130                1135                1140


Arg Thr  Asp Val Asp Phe Leu  Ile Ser Pro Val Lys  Asn Ser Asp
    1145                1150                1155


Gly Ile  Phe Tyr Asp Ser Arg  Asn Tyr Glu Ala Gln  Glu Asn Ala
    1160                1165                1170


Ile Leu  Pro Lys Asn Ala Asp  Ala Asn Gly Ala Tyr  Asn Ile Ala
    1175                1180                1185


Arg Lys  Val Leu Trp Ala Ile  Gly Gln Phe Lys Lys  Ala Glu Asp
    1190                1195                1200


Glu Lys  Leu Asp Lys Val Lys  Ile Ala Ser Asn Lys  Glu Trp Leu
    1205                1210                1215


Glu Tyr  Ala Gln Thr Ser Val  Lys His
    1220                1225


<210>  12
<211>  1264
<212>  PRT
<213>  Leptospira inadai

<400>  12

Met Glu Asp Tyr Ser Gly Phe Val Asn Ile Tyr Ser Ile Gln Lys Thr
1               5                   10                  15
```

```
Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu His Ile Glu
            20                  25                  30

Lys Lys Gly Phe Leu Lys Lys Asp Lys Ile Arg Ala Glu Asp Tyr Lys
        35                  40                  45

Ala Val Lys Lys Ile Ile Asp Lys Tyr His Arg Ala Tyr Ile Glu Glu
    50                  55                  60

Val Phe Asp Ser Val Leu His Gln Lys Lys Lys Lys Asp Lys Thr Arg
65                  70                  75                  80

Phe Ser Thr Gln Phe Ile Lys Glu Ile Lys Glu Phe Ser Glu Leu Tyr
                85                  90                  95

Tyr Lys Thr Glu Lys Asn Ile Pro Asp Lys Glu Arg Leu Glu Ala Leu
            100                 105                 110

Ser Glu Lys Leu Arg Lys Met Leu Val Gly Ala Phe Lys Gly Glu Phe
        115                 120                 125

Ser Glu Glu Val Ala Glu Lys Tyr Asn Lys Asn Leu Phe Ser Lys Glu
    130                 135                 140

Leu Ile Arg Asn Glu Ile Glu Lys Phe Cys Glu Thr Asp Glu Glu Arg
145                 150                 155                 160

Lys Gln Val Ser Asn Phe Lys Ser Phe Thr Thr Tyr Phe Thr Gly Phe
                165                 170                 175

His Ser Asn Arg Gln Asn Ile Tyr Ser Asp Glu Lys Lys Ser Thr Ala
            180                 185                 190

Ile Gly Tyr Arg Ile Ile His Gln Asn Leu Pro Lys Phe Leu Asp Asn
        195                 200                 205

Leu Lys Ile Ile Glu Ser Ile Gln Arg Arg Phe Lys Asp Phe Pro Trp
    210                 215                 220

Ser Asp Leu Lys Lys Asn Leu Lys Lys Ile Asp Lys Asn Ile Lys Leu
225                 230                 235                 240

Thr Glu Tyr Phe Ser Ile Asp Gly Phe Val Asn Val Leu Asn Gln Lys
                245                 250                 255

Gly Ile Asp Ala Tyr Asn Thr Ile Leu Gly Gly Lys Ser Glu Glu Ser
            260                 265                 270

Gly Glu Lys Ile Gln Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Arg Gln
        275                 280                 285
```

```
Lys Asn Asn Ile Asp Arg Lys Asn Pro Leu Asn Val Lys Ile Leu Phe
    290                 295                 300

Lys Gln Ile Leu Gly Asp Arg Glu Thr Lys Ser Phe Ile Pro Glu Ala
305                 310                 315                 320

Phe Pro Asp Asp Gln Ser Val Leu Asn Ser Ile Thr Glu Phe Ala Lys
                325                 330                 335

Tyr Leu Lys Leu Asp Lys Lys Lys Lys Ser Ile Ile Ala Glu Leu Lys
            340                 345                 350

Lys Phe Leu Ser Ser Phe Asn Arg Tyr Glu Leu Asp Gly Ile Tyr Leu
        355                 360                 365

Ala Asn Asp Asn Ser Leu Ala Ser Ile Ser Thr Phe Leu Phe Asp Asp
    370                 375                 380

Trp Ser Phe Ile Lys Lys Ser Val Ser Phe Lys Tyr Asp Glu Ser Val
385                 390                 395                 400

Gly Asp Pro Lys Lys Lys Ile Lys Ser Pro Leu Lys Tyr Glu Lys Glu
                405                 410                 415

Lys Glu Lys Trp Leu Lys Gln Lys Tyr Tyr Thr Ile Ser Phe Leu Asn
            420                 425                 430

Asp Ala Ile Glu Ser Tyr Ser Lys Ser Gln Asp Glu Lys Arg Val Lys
        435                 440                 445

Ile Arg Leu Glu Ala Tyr Phe Ala Glu Phe Lys Ser Lys Asp Asp Ala
    450                 455                 460

Lys Lys Gln Phe Asp Leu Leu Glu Arg Ile Glu Glu Ala Tyr Ala Ile
465                 470                 475                 480

Val Glu Pro Leu Leu Gly Ala Glu Tyr Pro Arg Asp Arg Asn Leu Lys
                485                 490                 495

Ala Asp Lys Lys Glu Val Gly Lys Ile Lys Asp Phe Leu Asp Ser Ile
            500                 505                 510

Lys Ser Leu Gln Phe Phe Leu Lys Pro Leu Leu Ser Ala Glu Ile Phe
        515                 520                 525

Asp Glu Lys Asp Leu Gly Phe Tyr Asn Gln Leu Glu Gly Tyr Tyr Glu
    530                 535                 540

Glu Ile Asp Ile Ser Gly His Leu Tyr Asn Lys Val Arg Asn Tyr Leu
545                 550                 555                 560
```

```
Thr Gly Lys Ile Tyr Ser Lys Glu Lys Phe Lys Leu Asn Phe Glu Asn
                565                 570                 575

Ser Thr Leu Leu Lys Gly Trp Asp Glu Asn Arg Glu Val Ala Asn Leu
            580                 585                 590

Cys Val Ile Phe Arg Glu Asp Gln Lys Tyr Tyr Leu Gly Val Met Asp
        595                 600                 605

Lys Glu Asn Asn Thr Ile Leu Ser Asp Ile Pro Lys Val Lys Pro Asn
    610                 615                 620

Glu Leu Phe Tyr Glu Lys Met Val Tyr Lys Leu Ile Pro Thr Pro His
625                 630                 635                 640

Met Gln Leu Pro Arg Ile Ile Phe Ser Ser Asp Asn Leu Ser Ile Tyr
                645                 650                 655

Asn Pro Ser Lys Ser Ile Leu Lys Ile Arg Glu Ala Lys Ser Phe Lys
            660                 665                 670

Glu Gly Lys Asn Phe Lys Leu Lys Asp Cys His Lys Phe Ile Asp Phe
        675                 680                 685

Tyr Lys Glu Ser Ile Ser Lys Asn Glu Asp Trp Ser Arg Phe Asp Phe
    690                 695                 700

Lys Phe Ser Lys Thr Ser Ser Tyr Glu Asn Ile Ser Glu Phe Tyr Arg
705                 710                 715                 720

Glu Val Glu Arg Gln Gly Tyr Asn Leu Asp Phe Lys Lys Val Ser Lys
                725                 730                 735

Phe Tyr Ile Asp Ser Leu Val Glu Asp Gly Lys Leu Tyr Leu Phe Gln
            740                 745                 750

Ile Tyr Asn Lys Asp Phe Ser Ile Phe Ser Lys Gly Lys Pro Asn Leu
        755                 760                 765

His Thr Ile Tyr Phe Arg Ser Leu Phe Ser Lys Glu Asn Leu Lys Asp
    770                 775                 780

Val Cys Leu Lys Leu Asn Gly Glu Ala Glu Met Phe Phe Arg Lys Lys
785                 790                 795                 800

Ser Ile Asn Tyr Asp Glu Lys Lys Lys Arg Glu Gly His His Pro Glu
                805                 810                 815

Leu Phe Glu Lys Leu Lys Tyr Pro Ile Leu Lys Asp Lys Arg Tyr Ser
            820                 825                 830

Glu Asp Lys Phe Gln Phe His Leu Pro Ile Ser Leu Asn Phe Lys Ser
```

```
        835                 840                 845

Lys Glu Arg Leu Asn Phe Asn Leu Lys Val Asn Glu Phe Leu Lys Arg
    850                 855                 860

Asn Lys Asp Ile Asn Ile Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu
865                 870                 875                 880

Leu Tyr Leu Val Met Ile Asn Gln Lys Gly Glu Ile Leu Lys Gln Thr
                885                 890                 895

Leu Leu Asp Ser Met Gln Ser Gly Lys Gly Arg Pro Glu Ile Asn Tyr
            900                 905                 910

Lys Glu Lys Leu Gln Glu Lys Glu Ile Glu Arg Asp Lys Ala Arg Lys
        915                 920                 925

Ser Trp Gly Thr Val Glu Asn Ile Lys Glu Leu Lys Glu Gly Tyr Leu
    930                 935                 940

Ser Ile Val Ile His Gln Ile Ser Lys Leu Met Val Glu Asn Asn Ala
945                 950                 955                 960

Ile Val Val Leu Glu Asp Leu Asn Ile Gly Phe Lys Arg Gly Arg Gln
                965                 970                 975

Lys Val Glu Arg Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp
            980                 985                 990

Lys Leu Asn Phe Leu Val Phe Lys  Glu Asn Lys Pro Thr  Glu Pro Gly
        995                 1000                1005

Gly Val  Leu Lys Ala Tyr Gln  Leu Thr Asp Glu Phe  Gln Ser Phe
    1010                1015                1020

Glu Lys  Leu Ser Lys Gln Thr  Gly Phe Leu Phe Tyr  Val Pro Ser
    1025                1030                1035

Trp Asn  Thr Ser Lys Ile Asp  Pro Arg Thr Gly Phe  Ile Asp Phe
    1040                1045                1050

Leu His  Pro Ala Tyr Glu Asn  Ile Glu Lys Ala Lys  Gln Trp Ile
    1055                1060                1065

Asn Lys  Phe Asp Ser Ile Arg  Phe Asn Ser Lys Met  Asp Trp Phe
    1070                1075                1080

Glu Phe  Thr Ala Asp Thr Arg  Lys Phe Ser Glu Asn  Leu Met Leu
    1085                1090                1095

Gly Lys  Asn Arg Val Trp Val  Ile Cys Thr Thr Asn  Val Glu Arg
    1100                1105                1110
```

```
Tyr Phe  Thr Ser Lys Thr Ala  Asn Ser Ser Ile Gln  Tyr Asn Ser
    1115                1120                1125


Ile Gln  Ile Thr Glu Lys Leu  Lys Glu Leu Phe Val  Asp Ile Pro
    1130                1135                1140


Phe Ser  Asn Gly Gln Asp Leu  Lys Pro Glu Ile Leu  Arg Lys Asn
    1145                1150                1155


Asp Ala  Val Phe Phe Lys Ser  Leu Leu Phe Tyr Ile  Lys Thr Thr
    1160                1165                1170


Leu Ser  Leu Arg Gln Asn Asn  Gly Lys Lys Gly Glu  Glu Glu Lys
    1175                1180                1185


Asp Phe  Ile Leu Ser Pro Val  Val Asp Ser Lys Gly  Arg Phe Phe
    1190                1195                1200


Asn Ser  Leu Glu Ala Ser Asp  Asp Glu Pro Lys Asp  Ala Asp Ala
    1205                1210                1215


Asn Gly  Ala Tyr His Ile Ala  Leu Lys Gly Leu Met  Asn Leu Leu
    1220                1225                1230


Val Leu  Asn Glu Thr Lys Glu  Glu Asn Leu Ser Arg  Pro Lys Trp
    1235                1240                1245


Lys Ile  Lys Asn Lys Asp Trp  Leu Glu Phe Val Trp  Glu Arg Asn
    1250                1255                1260


Arg
```

<210> 13
<211> 1373
<212> PRT
<213> Moraxella bovoculi

<400> 13

```
Met Leu Phe Gln Asp Phe Thr His Leu Tyr Pro Leu Ser Lys Thr Val
1               5                   10                  15


Arg Phe Glu Leu Phe Ile Asp Arg Thr Leu Glu His Ile His Ala Lys
            20                  25                  30


Asn Phe Leu Ser Gln Asp Glu Thr Met Ala Asp Met His Gln Lys Val
        35                  40                  45


Lys Val Ile Leu Asp Asp Tyr His Arg Asp Phe Ile Ala Asp Met Met
    50                  55                  60
```

```
Gly Glu Val Lys Leu Thr Lys Leu Ala Glu Phe Tyr Asp Val Tyr Leu
65                  70                  75                  80

Lys Phe Arg Lys Asn Pro Lys Asp Asp Glu Leu Gln Lys Ala Gln Leu
                85                  90                  95

Lys Asp Leu Gln Ala Val Leu Arg Lys Glu Ile Val Lys Pro Ile Gly
            100                 105                 110

Asn Gly Gly Lys Tyr Lys Ala Gly Tyr Asp Arg Leu Phe Gly Ala Lys
        115                 120                 125

Leu Phe Lys Asp Gly Lys Glu Leu Gly Asp Leu Ala Lys Phe Val Ile
    130                 135                 140

Ala Gln Glu Gly Glu Ser Ser Pro Lys Leu Ala His Leu Ala His Phe
145                 150                 155                 160

Glu Lys Phe Ser Thr Tyr Phe Thr Gly Phe His Asp Asn Arg Lys Asn
                165                 170                 175

Met Tyr Ser Asp Glu Asp Lys His Thr Ala Ile Ala Tyr Arg Leu Ile
            180                 185                 190

His Glu Asn Leu Pro Arg Phe Ile Asp Asn Leu Gln Ile Leu Thr Thr
        195                 200                 205

Ile Lys Gln Lys His Ser Ala Leu Tyr Asp Gln Ile Ile Asn Glu Leu
    210                 215                 220

Thr Ala Ser Gly Leu Asp Val Ser Leu Ala Ser His Leu Asp Gly Tyr
225                 230                 235                 240

His Lys Leu Leu Thr Gln Glu Gly Ile Thr Ala Tyr Asn Thr Leu Leu
                245                 250                 255

Gly Gly Ile Ser Gly Glu Ala Gly Ser Pro Lys Ile Gln Gly Ile Asn
            260                 265                 270

Glu Leu Ile Asn Ser His His Asn Gln His Cys His Lys Ser Glu Arg
        275                 280                 285

Ile Ala Lys Leu Arg Pro Leu His Lys Gln Ile Leu Ser Asp Gly Met
    290                 295                 300

Ser Val Ser Phe Leu Pro Ser Lys Phe Ala Asp Asp Ser Glu Met Cys
305                 310                 315                 320

Gln Ala Val Asn Glu Phe Tyr Arg His Tyr Ala Asp Val Phe Ala Lys
                325                 330                 335

Val Gln Ser Leu Phe Asp Gly Phe Asp Asp His Gln Lys Asp Gly Ile
```

```
            340                 345                 350

Tyr Val Glu His Lys Asn Leu Asn Glu Leu Ser Lys Gln Ala Phe Gly
        355                 360                 365

Asp Phe Ala Leu Leu Gly Arg Val Leu Asp Gly Tyr Tyr Val Asp Val
    370                 375                 380

Val Asn Pro Glu Phe Asn Glu Arg Phe Ala Lys Ala Lys Thr Asp Asn
385                 390                 395                 400

Ala Lys Ala Lys Leu Thr Lys Glu Lys Asp Lys Phe Ile Lys Gly Val
                405                 410                 415

His Ser Leu Ala Ser Leu Glu Gln Ala Ile Glu His Tyr Thr Ala Arg
            420                 425                 430

His Asp Asp Glu Ser Val Gln Ala Gly Lys Leu Gly Gln Tyr Phe Lys
        435                 440                 445

His Gly Leu Ala Gly Val Asp Asn Pro Ile Gln Lys Ile His Asn Asn
    450                 455                 460

His Ser Thr Ile Lys Gly Phe Leu Glu Arg Glu Arg Pro Ala Gly Glu
465                 470                 475                 480

Arg Ala Leu Pro Lys Ile Lys Ser Gly Lys Asn Pro Glu Met Thr Gln
                485                 490                 495

Leu Arg Gln Leu Lys Glu Leu Leu Asp Asn Ala Leu Asn Val Ala His
            500                 505                 510

Phe Ala Lys Leu Leu Thr Thr Lys Thr Thr Leu Asp Asn Gln Asp Gly
        515                 520                 525

Asn Phe Tyr Gly Glu Phe Gly Val Leu Tyr Asp Glu Leu Ala Lys Ile
    530                 535                 540

Pro Thr Leu Tyr Asn Lys Val Arg Asp Tyr Leu Ser Gln Lys Pro Phe
545                 550                 555                 560

Ser Thr Glu Lys Tyr Lys Leu Asn Phe Gly Asn Pro Thr Leu Leu Asn
                565                 570                 575

Gly Trp Asp Leu Asn Lys Glu Lys Asp Asn Phe Gly Val Ile Leu Gln
            580                 585                 590

Lys Asp Gly Cys Tyr Tyr Leu Ala Leu Leu Asp Lys Ala His Lys Lys
        595                 600                 605

Val Phe Asp Asn Ala Pro Asn Thr Gly Lys Ser Ile Tyr Gln Lys Met
    610                 615                 620
```

```
Ile Tyr Lys Tyr Leu Glu Val Arg Lys Gln Phe Pro Lys Val Phe Phe
625                 630                 635                 640

Ser Lys Glu Ala Ile Ala Ile Asn Tyr His Pro Ser Lys Glu Leu Val
                645                 650                 655

Glu Ile Lys Asp Lys Gly Arg Gln Arg Ser Asp Asp Glu Arg Leu Lys
            660                 665                 670

Leu Tyr Arg Phe Ile Leu Glu Cys Leu Lys Ile His Pro Lys Tyr Asp
        675                 680                 685

Lys Lys Phe Glu Gly Ala Ile Gly Asp Ile Gln Leu Phe Lys Lys Asp
    690                 695                 700

Lys Lys Gly Arg Glu Val Pro Ile Ser Glu Lys Asp Leu Phe Lys Asp
705                 710                 715                 720

Ile Asn Gly Ile Phe Ser Ser Lys Pro Lys Leu Glu Met Glu Asp Phe
                725                 730                 735

Phe Ile Gly Glu Phe Lys Arg Tyr Asn Pro Ser Gln Asp Leu Val Asp
            740                 745                 750

Gln Tyr Asn Ile Tyr Lys Lys Ile Asp Ser Asn Asp Asn Arg Lys Lys
        755                 760                 765

Glu Asn Phe Tyr Asn Asn His Pro Lys Phe Lys Lys Asp Leu Val Arg
    770                 775                 780

Tyr Tyr Tyr Glu Ser Met Cys Lys His Glu Glu Trp Glu Glu Ser Phe
785                 790                 795                 800

Glu Phe Ser Lys Lys Leu Gln Asp Ile Gly Cys Tyr Val Asp Val Asn
                805                 810                 815

Glu Leu Phe Thr Glu Ile Glu Thr Arg Arg Leu Asn Tyr Lys Ile Ser
            820                 825                 830

Phe Cys Asn Ile Asn Ala Asp Tyr Ile Asp Glu Leu Val Glu Gln Gly
        835                 840                 845

Gln Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Pro Lys Ala
    850                 855                 860

His Gly Lys Pro Asn Leu His Thr Leu Tyr Phe Lys Ala Leu Phe Ser
865                 870                 875                 880

Glu Asp Asn Leu Ala Asp Pro Ile Tyr Lys Leu Asn Gly Glu Ala Gln
                885                 890                 895
```

```
Ile Phe Tyr Arg Lys Ala Ser Leu Asp Met Asn Glu Thr Thr Ile His
            900                 905                 910

Arg Ala Gly Glu Val Leu Glu Asn Lys Asn Pro Asp Asn Pro Lys Lys
        915                 920                 925

Arg Gln Phe Val Tyr Asp Ile Ile Lys Asp Lys Arg Tyr Thr Gln Lys
    930                 935                 940

Asp Phe Met Leu His Val Pro Ile Thr Met Asn Phe Gly Val Gln Gly
945                 950                 955                 960

Met Thr Ile Lys Glu Phe Asn Lys Lys Val Asn Gln Ser Ile Gln Gln
                965                 970                 975

Tyr Asp Glu Val Asn Val Ile Gly Ile Asp Arg Gly Glu Arg His Leu
            980                 985                 990

Leu Tyr Leu Thr Val Ile Asn Ser  Lys Gly Glu Ile Leu  Glu Gln Cys
        995                 1000                1005

Ser Leu  Asn Asp Ile Thr Thr  Ala Ser Ala Asn Gly  Thr Gln Met
    1010                1015                1020

Thr Thr  Pro Tyr His Lys Ile  Leu Asp Lys Arg Glu  Ile Glu Arg
    1025                1030                1035

Leu Asn  Ala Arg Val Gly Trp  Gly Glu Ile Glu Thr  Ile Lys Glu
    1040                1045                1050

Leu Lys  Ser Gly Tyr Leu Ser  His Val Val His Gln  Ile Ser Gln
    1055                1060                1065

Leu Met  Leu Lys Tyr Asn Ala  Ile Val Val Leu Glu  Asp Leu Asn
    1070                1075                1080

Phe Gly  Phe Lys Arg Gly Arg  Phe Lys Val Glu Lys  Gln Ile Tyr
    1085                1090                1095

Gln Asn  Phe Glu Asn Ala Leu  Ile Lys Lys Leu Asn  His Leu Val
    1100                1105                1110

Leu Lys  Asp Lys Ala Asp Asp  Glu Ile Gly Ser Tyr  Lys Asn Ala
    1115                1120                1125

Leu Gln  Leu Thr Asn Asn Phe  Thr Asp Leu Lys Ser  Ile Gly Lys
    1130                1135                1140

Gln Thr  Gly Phe Leu Phe Tyr  Val Pro Ala Trp Asn  Thr Ser Lys
    1145                1150                1155
```

```
Ile Asp  Pro Glu Thr Gly Phe  Val Asp Leu Leu Lys  Pro Arg Tyr
    1160                 1165                 1170


Glu Asn  Ile Gln Ala Ser Gln  Ala Phe Phe Gly Lys  Phe Asp Lys
    1175                 1180                 1185


Ile Cys  Tyr Asn Ala Asp Lys  Asp Tyr Phe Glu Phe  His Ile Asp
    1190                 1195                 1200


Tyr Ala  Lys Phe Thr Asp Lys  Ala Lys Asn Ser Arg  Gln Ile Trp
    1205                 1210                 1215


Thr Ile  Cys Ser His Gly Asp  Lys Arg Tyr Val Tyr  Asp Lys Thr
    1220                 1225                 1230


Ala Asn  Gln Asn Lys Gly Ala  Ala Lys Gly Ile Asn  Val Asn Asp
    1235                 1240                 1245


Ile Leu  Lys Ser Leu Phe Ala  Arg His His Ile Asn  Glu Lys Gln
    1250                 1255                 1260


Pro Asn  Leu Val Met Asp Ile  Cys Gln Asn Asn Asp  Lys Glu Phe
    1265                 1270                 1275


His Lys  Ser Leu Met Tyr Leu  Leu Lys Thr Leu Leu  Ala Leu Arg
    1280                 1285                 1290


Tyr Ser  Asn Ala Ser Ser Asp  Glu Asp Phe Ile Leu  Ser Pro Val
    1295                 1300                 1305


Ala Asn  Asp Glu Gly Val Phe  Phe Asn Ser Ala Leu  Ala Asp Asp
    1310                 1315                 1320


Thr Gln  Pro Gln Asn Ala Asp  Ala Asn Gly Ala Tyr  His Ile Ala
    1325                 1330                 1335


Leu Lys  Gly Leu Trp Leu Leu  Asn Glu Leu Lys Asn  Ser Asp Asp
    1340                 1345                 1350


Leu Asn  Lys Val Lys Leu Ala  Ile Asp Asn Gln Thr  Trp Leu Asn
    1355                 1360                 1365


Phe Ala  Gln Asn Arg
    1370


<210>  14
<211>  1352
<212>  PRT
<213>  Unknown

<220>
<223>  Parcubacteria bacterium

<400>  14
```

```
Met Glu Asn Ile Phe Asp Gln Phe Ile Gly Lys Tyr Ser Leu Ser Lys
1               5                   10                  15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Glu Asp Phe Leu
            20                  25                  30

Lys Ile Asn Lys Val Phe Glu Lys Asp Gln Thr Ile Asp Asp Ser Tyr
        35                  40                  45

Asn Gln Ala Lys Phe Tyr Phe Asp Ser Leu His Gln Lys Phe Ile Asp
    50                  55                  60

Ala Ala Leu Ala Ser Asp Lys Thr Ser Glu Leu Ser Phe Gln Asn Phe
65                  70                  75                  80

Ala Asp Val Leu Glu Lys Gln Asn Lys Ile Ile Leu Asp Lys Lys Arg
                85                  90                  95

Glu Met Gly Ala Leu Arg Lys Arg Asp Lys Asn Ala Val Gly Ile Asp
            100                 105                 110

Arg Leu Gln Lys Glu Ile Asn Asp Ala Glu Asp Ile Ile Gln Lys Glu
        115                 120                 125

Lys Glu Lys Ile Tyr Lys Asp Val Arg Thr Leu Phe Asp Asn Glu Ala
    130                 135                 140

Glu Ser Trp Lys Thr Tyr Tyr Gln Glu Arg Glu Val Asp Gly Lys Lys
145                 150                 155                 160

Ile Thr Glu Ser Lys Ala Asp Leu Lys Gln Lys Gly Ala Asp Phe Leu
                165                 170                 175

Thr Ala Ala Gly Ile Leu Lys Val Leu Lys Tyr Glu Phe Pro Glu Glu
            180                 185                 190

Lys Glu Lys Glu Phe Gln Ala Lys Asn Gln Pro Ser Leu Phe Val Glu
        195                 200                 205

Glu Lys Glu Asn Pro Gly Gln Lys Arg Tyr Ile Phe Asp Ser Phe Asp
    210                 215                 220

Lys Phe Ala Gly Tyr Leu Thr Lys Phe Gln Gln Thr Lys Lys Asn Leu
225                 230                 235                 240

Tyr Ala Ala Asp Gly Thr Ser Thr Ala Val Ala Thr Arg Ile Ala Asp
                245                 250                 255

Asn Phe Ile Ile Phe His Gln Asn Thr Lys Val Phe Arg Asp Lys Tyr
            260                 265                 270
```

```
Lys Asn Asn His Thr Asp Leu Gly Phe Asp Glu Glu Asn Ile Phe Glu
        275                 280                 285

Ile Glu Arg Tyr Lys Asn Cys Leu Leu Gln Arg Glu Ile Glu His Ile
    290                 295                 300

Lys Asn Glu Asn Ser Tyr Asn Lys Ile Ile Gly Arg Ile Asn Lys Lys
305                 310                 315                 320

Ile Lys Glu Tyr Arg Asp Gln Lys Ala Lys Asp Thr Lys Leu Thr Lys
                325                 330                 335

Ser Asp Phe Pro Phe Phe Lys Asn Leu Asp Lys Gln Ile Leu Gly Glu
            340                 345                 350

Val Glu Lys Glu Lys Gln Leu Ile Glu Lys Thr Arg Glu Lys Thr Glu
        355                 360                 365

Glu Asp Val Leu Ile Glu Arg Phe Lys Glu Phe Ile Glu Asn Asn Glu
    370                 375                 380

Glu Arg Phe Thr Ala Ala Lys Lys Leu Met Asn Ala Phe Cys Asn Gly
385                 390                 395                 400

Glu Phe Glu Ser Glu Tyr Glu Gly Ile Tyr Leu Lys Asn Lys Ala Ile
                405                 410                 415

Asn Thr Ile Ser Arg Arg Trp Phe Val Ser Asp Arg Asp Phe Glu Leu
            420                 425                 430

Lys Leu Pro Gln Gln Lys Ser Lys Asn Lys Ser Glu Lys Asn Glu Pro
        435                 440                 445

Lys Val Lys Lys Phe Ile Ser Ile Ala Glu Ile Lys Asn Ala Val Glu
    450                 455                 460

Glu Leu Asp Gly Asp Ile Phe Lys Ala Val Phe Tyr Asp Lys Lys Ile
465                 470                 475                 480

Ile Ala Gln Gly Gly Ser Lys Leu Glu Gln Phe Leu Val Ile Trp Lys
                485                 490                 495

Tyr Glu Phe Glu Tyr Leu Phe Arg Asp Ile Glu Arg Glu Asn Gly Glu
            500                 505                 510

Lys Leu Leu Gly Tyr Asp Ser Cys Leu Lys Ile Ala Lys Gln Leu Gly
        515                 520                 525

Ile Phe Pro Gln Glu Lys Glu Ala Arg Glu Lys Ala Thr Ala Val Ile
    530                 535                 540

Lys Asn Tyr Ala Asp Ala Gly Leu Gly Ile Phe Gln Met Met Lys Tyr
```

```
545                 550                 555                 560

Phe Ser Leu Asp Asp Lys Asp Arg Lys Asn Thr Pro Gly Gln Leu Ser
                565                 570                 575

Thr Asn Phe Tyr Ala Glu Tyr Asp Gly Tyr Tyr Lys Asp Phe Glu Phe
            580                 585                 590

Ile Lys Tyr Tyr Asn Glu Phe Arg Asn Phe Ile Thr Lys Lys Pro Phe
        595                 600                 605

Asp Glu Asp Lys Ile Lys Leu Asn Phe Glu Asn Gly Ala Leu Leu Lys
    610                 615                 620

Gly Trp Asp Glu Asn Lys Glu Tyr Asp Phe Met Gly Val Ile Leu Lys
625                 630                 635                 640

Lys Glu Gly Arg Leu Tyr Leu Gly Ile Met His Lys Asn His Arg Lys
                645                 650                 655

Leu Phe Gln Ser Met Gly Asn Ala Lys Gly Asp Asn Ala Asn Arg Tyr
            660                 665                 670

Gln Lys Met Ile Tyr Lys Gln Ile Ala Asp Ala Ser Lys Asp Val Pro
        675                 680                 685

Arg Leu Leu Leu Thr Ser Lys Lys Ala Met Glu Lys Phe Lys Pro Ser
    690                 695                 700

Gln Glu Ile Leu Arg Ile Lys Lys Glu Lys Thr Phe Lys Arg Glu Ser
705                 710                 715                 720

Lys Asn Phe Ser Leu Arg Asp Leu His Ala Leu Ile Glu Tyr Tyr Arg
                725                 730                 735

Asn Cys Ile Pro Gln Tyr Ser Asn Trp Ser Phe Tyr Asp Phe Gln Phe
            740                 745                 750

Gln Asp Thr Gly Lys Tyr Gln Asn Ile Lys Glu Phe Thr Asp Asp Val
        755                 760                 765

Gln Lys Tyr Gly Tyr Lys Ile Ser Phe Arg Asp Ile Asp Asp Glu Tyr
    770                 775                 780

Ile Asn Gln Ala Leu Asn Glu Gly Lys Met Tyr Leu Phe Glu Val Val
785                 790                 795                 800

Asn Lys Asp Ile Tyr Asn Thr Lys Asn Gly Ser Lys Asn Leu His Thr
                805                 810                 815

Leu Tyr Phe Glu His Ile Leu Ser Ala Glu Asn Leu Asn Asp Pro Val
            820                 825                 830
```

```
Phe Lys Leu Ser Gly Met Ala Glu Ile Phe Gln Arg Gln Pro Ser Val
        835                 840                 845

Asn Glu Arg Glu Lys Ile Thr Thr Gln Lys Asn Gln Cys Ile Leu Asp
    850                 855                 860

Lys Gly Asp Arg Ala Tyr Lys Tyr Arg Arg Tyr Thr Glu Lys Lys Ile
865                 870                 875                 880

Met Phe His Met Ser Leu Val Leu Asn Thr Gly Lys Gly Glu Ile Lys
                885                 890                 895

Gln Val Gln Phe Asn Lys Ile Ile Asn Gln Arg Ile Ser Ser Ser Asp
            900                 905                 910

Asn Glu Met Arg Val Asn Val Ile Gly Ile Asp Arg Gly Glu Lys Asn
        915                 920                 925

Leu Leu Tyr Tyr Ser Val Val Lys Gln Asn Gly Glu Ile Ile Glu Gln
    930                 935                 940

Ala Ser Leu Asn Glu Ile Asn Gly Val Asn Tyr Arg Asp Lys Leu Ile
945                 950                 955                 960

Glu Arg Glu Lys Glu Arg Leu Lys Asn Arg Gln Ser Trp Lys Pro Val
                965                 970                 975

Val Lys Ile Lys Asp Leu Lys Lys Gly Tyr Ile Ser His Val Ile His
            980                 985                 990

Lys Ile Cys Gln Leu Ile Glu Lys  Tyr Ser Ala Ile Val  Val Leu Glu
        995                 1000                 1005

Asp Leu  Asn Met Arg Phe Lys  Gln Ile Arg Gly Gly  Ile Glu Arg
    1010                1015                1020

Ser Val  Tyr Gln Gln Phe Glu  Lys Ala Leu Ile Asp  Lys Leu Gly
    1025                1030                1035

Tyr Leu  Val Phe Lys Asp Asn  Arg Asp Leu Arg Ala  Pro Gly Gly
    1040                1045                1050

Val Leu  Asn Gly Tyr Gln Leu  Ser Ala Pro Phe Val  Ser Phe Glu
    1055                1060                1065

Lys Met  Arg Lys Gln Thr Gly  Ile Leu Phe Tyr Thr  Gln Ala Glu
    1070                1075                1080

Tyr Thr  Ser Lys Thr Asp Pro  Ile Thr Gly Phe Arg  Lys Asn Val
    1085                1090                1095
```

```
Tyr Ile  Ser Asn Ser Ala Ser  Leu Asp Lys Ile Lys  Glu Ala Val
    1100                 1105                 1110

Lys Lys  Phe Asp Ala Ile Gly  Trp Asp Gly Lys Glu  Gln Ser Tyr
    1115                 1120                 1125

Phe Phe  Lys Tyr Asn Pro Tyr  Asn Leu Ala Asp Glu  Lys Tyr Lys
    1130                 1135                 1140

Asn Ser  Thr Val Ser Lys Glu  Trp Ala Ile Phe Ala  Ser Ala Pro
    1145                 1150                 1155

Arg Ile  Arg Arg Gln Lys Gly  Glu Asp Gly Tyr Trp  Lys Tyr Asp
    1160                 1165                 1170

Arg Val  Lys Val Asn Glu Glu  Phe Glu Lys Leu Leu  Lys Val Trp
    1175                 1180                 1185

Asn Phe  Val Asn Pro Lys Ala  Thr Asp Ile Lys Gln  Glu Ile Ile
    1190                 1195                 1200

Lys Lys  Ile Lys Ala Gly Asp  Leu Gln Gly Glu Lys  Glu Leu Asp
    1205                 1210                 1215

Gly Arg  Leu Arg Asn Phe Trp  His Ser Phe Ile Tyr  Leu Phe Asn
    1220                 1225                 1230

Leu Val  Leu Glu Leu Arg Asn  Ser Phe Ser Leu Gln  Ile Lys Ile
    1235                 1240                 1245

Lys Ala  Gly Glu Val Ile Ala  Val Asp Glu Gly Val  Asp Phe Ile
    1250                 1255                 1260

Ala Ser  Pro Val Lys Pro Phe  Phe Thr Thr Pro Asn  Pro Tyr Ile
    1265                 1270                 1275

Pro Ser  Asn Leu Cys Trp Leu  Ala Val Glu Asn Ala  Asp Ala Asn
    1280                 1285                 1290

Gly Ala  Tyr Asn Ile Ala Arg  Lys Gly Val Met Ile  Leu Lys Lys
    1295                 1300                 1305

Ile Arg  Glu His Ala Lys Lys  Asp Pro Glu Phe Lys  Lys Leu Pro
    1310                 1315                 1320

Asn Leu  Phe Ile Ser Asn Ala  Glu Trp Asp Glu Ala  Ala Arg Asp
    1325                 1330                 1335

Trp Gly  Lys Tyr Ala Gly Thr  Thr Ala Leu Asn Leu  Asp His
    1340                 1345                 1350
```

<210> 15
<211> 1260
<212> PRT
<213> Porphyromonas crevioricanis

<400> 15

Met Asp Ser Leu Lys Asp Phe Thr Asn Leu Tyr Pro Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Lys Pro Val Gly Lys Thr Leu Glu Asn Ile Glu
20 25 30

Lys Ala Gly Ile Leu Lys Glu Asp Glu His Arg Ala Glu Ser Tyr Arg
35 40 45

Arg Val Lys Lys Ile Ile Asp Thr Tyr His Lys Val Phe Ile Asp Ser
50 55 60

Ser Leu Glu Asn Met Ala Lys Met Gly Ile Glu Asn Glu Ile Lys Ala
65 70 75 80

Met Leu Gln Ser Phe Cys Glu Leu Tyr Lys Lys Asp His Arg Thr Glu
85 90 95

Gly Glu Asp Lys Ala Leu Asp Lys Ile Arg Ala Val Leu Arg Gly Leu
100 105 110

Ile Val Gly Ala Phe Thr Gly Val Cys Gly Arg Arg Glu Asn Thr Val
115 120 125

Gln Asn Glu Lys Tyr Glu Ser Leu Phe Lys Glu Lys Leu Ile Lys Glu
130 135 140

Ile Leu Pro Asp Phe Val Leu Ser Thr Glu Ala Glu Ser Leu Pro Phe
145 150 155 160

Ser Val Glu Glu Ala Thr Arg Ser Leu Lys Glu Phe Asp Ser Phe Thr
165 170 175

Ser Tyr Phe Ala Gly Phe Tyr Glu Asn Arg Lys Asn Ile Tyr Ser Thr
180 185 190

Lys Pro Gln Ser Thr Ala Ile Ala Tyr Arg Leu Ile His Glu Asn Leu
195 200 205

Pro Lys Phe Ile Asp Asn Ile Leu Val Phe Gln Lys Ile Lys Glu Pro
210 215 220

Ile Ala Lys Glu Leu Glu His Ile Arg Ala Asp Phe Ser Ala Gly Gly
225 230 235 240

Tyr Ile Lys Lys Asp Glu Arg Leu Glu Asp Ile Phe Ser Leu Asn Tyr
245 250 255

```
Tyr Ile His Val Leu Ser Gln Ala Gly Ile Glu Lys Tyr Asn Ala Leu
            260                 265                 270

Ile Gly Lys Ile Val Thr Glu Gly Asp Gly Glu Met Lys Gly Leu Asn
        275                 280                 285

Glu His Ile Asn Leu Tyr Asn Gln Gln Arg Gly Arg Glu Asp Arg Leu
    290                 295                 300

Pro Leu Phe Arg Pro Leu Tyr Lys Gln Ile Leu Ser Asp Arg Glu Gln
305                 310                 315                 320

Leu Ser Tyr Leu Pro Glu Ser Phe Glu Lys Asp Glu Glu Leu Leu Arg
                325                 330                 335

Ala Leu Lys Glu Phe Tyr Asp His Ile Ala Glu Asp Ile Leu Gly Arg
            340                 345                 350

Thr Gln Gln Leu Met Thr Ser Ile Ser Glu Tyr Asp Leu Ser Arg Ile
        355                 360                 365

Tyr Val Arg Asn Asp Ser Gln Leu Thr Asp Ile Ser Lys Lys Met Leu
    370                 375                 380

Gly Asp Trp Asn Ala Ile Tyr Met Ala Arg Glu Arg Ala Tyr Asp His
385                 390                 395                 400

Glu Gln Ala Pro Lys Arg Ile Thr Ala Lys Tyr Glu Arg Asp Arg Ile
                405                 410                 415

Lys Ala Leu Lys Gly Glu Glu Ser Ile Ser Leu Ala Asn Leu Asn Ser
            420                 425                 430

Cys Ile Ala Phe Leu Asp Asn Val Arg Asp Cys Arg Val Asp Thr Tyr
        435                 440                 445

Leu Ser Thr Leu Gly Gln Lys Glu Gly Pro His Gly Leu Ser Asn Leu
    450                 455                 460

Val Glu Asn Val Phe Ala Ser Tyr His Glu Ala Glu Gln Leu Leu Ser
465                 470                 475                 480

Phe Pro Tyr Pro Glu Glu Asn Asn Leu Ile Gln Asp Lys Asp Asn Val
                485                 490                 495

Val Leu Ile Lys Asn Leu Leu Asp Asn Ile Ser Asp Leu Gln Arg Phe
            500                 505                 510

Leu Lys Pro Leu Trp Gly Met Gly Asp Glu Pro Asp Lys Asp Glu Arg
        515                 520                 525
```

```
Phe Tyr Gly Glu Tyr Asn Tyr Ile Arg Gly Ala Leu Asp Gln Val Ile
    530                 535                 540

Pro Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Arg Lys Pro Tyr Ser
545                 550                 555                 560

Thr Arg Lys Val Lys Leu Asn Phe Gly Asn Ser Gln Leu Leu Ser Gly
                565                 570                 575

Trp Asp Arg Asn Lys Glu Lys Asp Asn Ser Cys Val Ile Leu Arg Lys
            580                 585                 590

Gly Gln Asn Phe Tyr Leu Ala Ile Met Asn Asn Arg His Lys Arg Ser
        595                 600                 605

Phe Glu Asn Lys Met Leu Pro Glu Tyr Lys Glu Gly Glu Pro Tyr Phe
    610                 615                 620

Glu Lys Met Asp Tyr Lys Phe Leu Pro Asp Pro Asn Lys Met Leu Pro
625                 630                 635                 640

Lys Val Phe Leu Ser Lys Lys Gly Ile Glu Ile Tyr Lys Pro Ser Pro
                645                 650                 655

Lys Leu Leu Glu Gln Tyr Gly His Gly Thr His Lys Lys Gly Asp Thr
            660                 665                 670

Phe Ser Met Asp Asp Leu His Glu Leu Ile Asp Phe Phe Lys His Ser
        675                 680                 685

Ile Glu Ala His Glu Asp Trp Lys Gln Phe Gly Phe Lys Phe Ser Asp
    690                 695                 700

Thr Ala Thr Tyr Glu Asn Val Ser Ser Phe Tyr Arg Glu Val Glu Asp
705                 710                 715                 720

Gln Gly Tyr Lys Leu Ser Phe Arg Lys Val Ser Glu Ser Tyr Val Tyr
                725                 730                 735

Ser Leu Ile Asp Gln Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
            740                 745                 750

Asp Phe Ser Pro Cys Ser Lys Gly Thr Pro Asn Leu His Thr Leu Tyr
        755                 760                 765

Trp Arg Met Leu Phe Asp Glu Arg Asn Leu Ala Asp Val Ile Tyr Lys
    770                 775                 780

Leu Asp Gly Lys Ala Glu Ile Phe Phe Arg Glu Lys Ser Leu Lys Asn
785                 790                 795                 800
```

```
Asp His Pro Thr His Pro Ala Gly Lys Pro Ile Lys Lys Lys Ser Arg
                805                 810                 815

Gln Lys Lys Gly Glu Glu Ser Leu Phe Glu Tyr Asp Leu Val Lys Asp
            820                 825                 830

Arg Arg Tyr Thr Met Asp Lys Phe Gln Phe His Val Pro Ile Thr Met
        835                 840                 845

Asn Phe Lys Cys Ser Ala Gly Ser Lys Val Asn Asp Met Val Asn Ala
    850                 855                 860

His Ile Arg Glu Ala Lys Asp Met His Val Ile Gly Ile Asp Arg Gly
865                 870                 875                 880

Glu Arg Asn Leu Leu Tyr Ile Cys Val Ile Asp Ser Arg Gly Thr Ile
                885                 890                 895

Leu Asp Gln Ile Ser Leu Asn Thr Ile Asn Asp Ile Asp Tyr His Asp
            900                 905                 910

Leu Leu Glu Ser Arg Asp Lys Asp Arg Gln Gln Glu His Arg Asn Trp
        915                 920                 925

Gln Thr Ile Glu Gly Ile Lys Glu Leu Lys Gln Gly Tyr Leu Ser Gln
    930                 935                 940

Ala Val His Arg Ile Ala Glu Leu Met Val Ala Tyr Lys Ala Val Val
945                 950                 955                 960

Ala Leu Glu Asp Leu Asn Met Gly Phe Lys Arg Gly Arg Gln Lys Val
                965                 970                 975

Glu Ser Ser Val Tyr Gln Gln Phe Glu Lys Gln Leu Ile Asp Lys Leu
            980                 985                 990

Asn Tyr Leu Val Asp Lys Lys Lys  Arg Pro Glu Asp Ile  Gly Gly Leu
        995                 1000                1005

Leu Arg  Ala Tyr Gln Phe Thr  Ala Pro Phe Lys Ser  Phe Lys Glu
    1010                1015                1020

Met Gly  Lys Gln Asn Gly Phe  Leu Phe Tyr Ile Pro  Ala Trp Asn
    1025                1030                1035

Thr Ser  Asn Ile Asp Pro Thr  Thr Gly Phe Val Asn  Leu Phe His
    1040                1045                1050

Val Gln  Tyr Glu Asn Val Asp  Lys Ala Lys Ser Phe  Phe Gln Lys
    1055                1060                1065

Phe Asp  Ser Ile Ser Tyr Asn  Pro Lys Lys Asp Trp  Phe Glu Phe
```

```
    1070                1075                1080

Ala Phe  Asp Tyr Lys Asn Phe  Thr Lys Lys Ala Glu  Gly Ser Arg
    1085                1090                1095

Ser Met  Trp Ile Leu Cys Thr  His Gly Ser Arg Ile  Lys Asn Phe
    1100                1105                1110

Arg Asn  Ser Gln Lys Asn Gly  Gln Trp Asp Ser Glu  Glu Phe Ala
    1115                1120                1125

Leu Thr  Glu Ala Phe Lys Ser  Leu Phe Val Arg Tyr  Glu Ile Asp
    1130                1135                1140

Tyr Thr  Ala Asp Leu Lys Thr  Ala Ile Val Asp Glu  Lys Gln Lys
    1145                1150                1155

Asp Phe  Phe Val Asp Leu Leu  Lys Leu Phe Lys Leu  Thr Val Gln
    1160                1165                1170

Met Arg  Asn Ser Trp Lys Glu  Lys Asp Leu Asp Tyr  Leu Ile Ser
    1175                1180                1185

Pro Val  Ala Gly Ala Asp Gly  Arg Phe Phe Asp Thr  Arg Glu Gly
    1190                1195                1200

Asn Lys  Ser Leu Pro Lys Asp  Ala Asp Ala Asn Gly  Ala Tyr Asn
    1205                1210                1215

Ile Ala  Leu Lys Gly Leu Trp  Ala Leu Arg Gln Ile  Arg Gln Thr
    1220                1225                1230

Ser Glu  Gly Gly Lys Leu Lys  Leu Ala Ile Ser Asn  Lys Glu Trp
    1235                1240                1245

Leu Gln  Phe Val Gln Glu Arg  Ser Tyr Glu Lys Asp
    1250                1255                1260


<210>  16
<211>  1324
<212>  PRT
<213>  Prevotella disiens

<400>  16

Met Glu Asn Tyr Gln Glu Phe Thr Asn Leu Phe Gln Leu Asn Lys Thr
1               5                  10                  15

Leu Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Cys Glu Leu Leu Glu
            20                  25                  30

Glu Gly Lys Ile Phe Ala Ser Gly Ser Phe Leu Glu Lys Asp Lys Val
        35                  40                  45
```

```
Arg Ala Asp Asn Val Ser Tyr Val Lys Lys Glu Ile Asp Lys Lys His
    50                  55                  60

Lys Ile Phe Ile Glu Glu Thr Leu Ser Ser Phe Ser Ile Ser Asn Asp
65                  70                  75                  80

Leu Leu Lys Gln Tyr Phe Asp Cys Tyr Asn Glu Leu Lys Ala Phe Lys
                85                  90                  95

Lys Asp Cys Lys Ser Asp Glu Glu Glu Val Lys Lys Thr Ala Leu Arg
            100                 105                 110

Asn Lys Cys Thr Ser Ile Gln Arg Ala Met Arg Glu Ala Ile Ser Gln
        115                 120                 125

Ala Phe Leu Lys Ser Pro Gln Lys Lys Leu Leu Ala Ile Lys Asn Leu
    130                 135                 140

Ile Glu Asn Val Phe Lys Ala Asp Glu Asn Val Gln His Phe Ser Glu
145                 150                 155                 160

Phe Thr Ser Tyr Phe Ser Gly Phe Glu Thr Asn Arg Glu Asn Phe Tyr
                165                 170                 175

Ser Asp Glu Glu Lys Ser Thr Ser Ile Ala Tyr Arg Leu Val His Asp
            180                 185                 190

Asn Leu Pro Ile Phe Ile Lys Asn Ile Tyr Ile Phe Glu Lys Leu Lys
        195                 200                 205

Glu Gln Phe Asp Ala Lys Thr Leu Ser Glu Ile Phe Glu Asn Tyr Lys
    210                 215                 220

Leu Tyr Val Ala Gly Ser Ser Leu Asp Glu Val Phe Ser Leu Glu Tyr
225                 230                 235                 240

Phe Asn Asn Thr Leu Thr Gln Lys Gly Ile Asp Asn Tyr Asn Ala Val
                245                 250                 255

Ile Gly Lys Ile Val Lys Glu Asp Lys Gln Glu Ile Gln Gly Leu Asn
            260                 265                 270

Glu His Ile Asn Leu Tyr Asn Gln Lys His Lys Asp Arg Arg Leu Pro
        275                 280                 285

Phe Phe Ile Ser Leu Lys Lys Gln Ile Leu Ser Asp Arg Glu Ala Leu
    290                 295                 300

Ser Trp Leu Pro Asp Met Phe Lys Asn Asp Ser Glu Val Ile Asp Ala
305                 310                 315                 320
```

```
Leu Lys Gly Phe Tyr Ile Glu Asp Gly Phe Glu Asn Asn Val Leu Thr
                325                 330                 335

Pro Leu Ala Thr Leu Leu Ser Ser Leu Asp Lys Tyr Asn Leu Asn Gly
            340                 345                 350

Ile Phe Ile Arg Asn Asn Glu Ala Leu Ser Ser Leu Ser Gln Asn Val
        355                 360                 365

Tyr Arg Asn Phe Ser Ile Asp Glu Ala Ile Asp Ala Gln Asn Ala Glu
    370                 375                 380

Leu Gln Thr Phe Asn Asn Tyr Glu Leu Ile Ala Asn Ala Leu Arg Ala
385                 390                 395                 400

Lys Ile Lys Lys Glu Thr Lys Gln Gly Arg Lys Ser Phe Glu Lys Tyr
                405                 410                 415

Glu Glu Tyr Ile Asp Lys Lys Val Lys Ala Ile Asp Ser Leu Ser Ile
            420                 425                 430

Gln Glu Ile Asn Glu Leu Val Glu Asn Tyr Val Ser Glu Phe Asn Ser
        435                 440                 445

Asn Ser Gly Asn Met Pro Arg Lys Val Glu Asp Tyr Phe Ser Leu Met
    450                 455                 460

Arg Lys Gly Asp Phe Gly Ser Asn Asp Leu Ile Glu Asn Ile Lys Thr
465                 470                 475                 480

Lys Leu Ser Ala Ala Glu Lys Leu Leu Gly Thr Lys Tyr Gln Glu Thr
                485                 490                 495

Ala Lys Asp Ile Phe Lys Lys Asp Glu Asn Ser Lys Leu Ile Lys Glu
            500                 505                 510

Leu Leu Asp Ala Thr Lys Gln Phe Gln His Phe Ile Lys Pro Leu Leu
        515                 520                 525

Gly Thr Gly Glu Glu Ala Asp Arg Asp Leu Val Phe Tyr Gly Asp Phe
    530                 535                 540

Leu Pro Leu Tyr Glu Lys Phe Glu Glu Leu Thr Leu Leu Tyr Asn Lys
545                 550                 555                 560

Val Arg Asn Arg Leu Thr Gln Lys Pro Tyr Ser Lys Asp Lys Ile Arg
                565                 570                 575

Leu Cys Phe Asn Lys Pro Lys Leu Met Thr Gly Trp Val Asp Ser Lys
            580                 585                 590

Thr Glu Lys Ser Asp Asn Gly Thr Gln Tyr Gly Gly Tyr Leu Phe Arg
```

```
        595                 600                 605

Lys Lys Asn Glu Ile Gly Glu Tyr Asp Tyr Phe Leu Gly Ile Ser Ser
    610                 615                 620

Lys Ala Gln Leu Phe Arg Lys Asn Glu Ala Val Ile Gly Asp Tyr Glu
625                 630                 635                 640

Arg Leu Asp Tyr Tyr Gln Pro Lys Ala Asn Thr Ile Tyr Gly Ser Ala
                645                 650                 655

Tyr Glu Gly Glu Asn Ser Tyr Lys Glu Asp Lys Lys Arg Leu Asn Lys
            660                 665                 670

Val Ile Ile Ala Tyr Ile Glu Gln Ile Lys Gln Thr Asn Ile Lys Lys
        675                 680                 685

Ser Ile Ile Glu Ser Ile Ser Lys Tyr Pro Asn Ile Ser Asp Asp Asp
    690                 695                 700

Lys Val Thr Pro Ser Ser Leu Leu Glu Lys Ile Lys Lys Val Ser Ile
705                 710                 715                 720

Asp Ser Tyr Asn Gly Ile Leu Ser Phe Lys Ser Phe Gln Ser Val Asn
                725                 730                 735

Lys Glu Val Ile Asp Asn Leu Leu Lys Thr Ile Ser Pro Leu Lys Asn
            740                 745                 750

Lys Ala Glu Phe Leu Asp Leu Ile Asn Lys Asp Tyr Gln Ile Phe Thr
        755                 760                 765

Glu Val Gln Ala Val Ile Asp Glu Ile Cys Lys Gln Lys Thr Phe Ile
    770                 775                 780

Tyr Phe Pro Ile Ser Asn Val Glu Leu Glu Lys Glu Met Gly Asp Lys
785                 790                 795                 800

Asp Lys Pro Leu Cys Leu Phe Gln Ile Ser Asn Lys Asp Leu Ser Phe
                805                 810                 815

Ala Lys Thr Phe Ser Ala Asn Leu Arg Lys Lys Arg Gly Ala Glu Asn
            820                 825                 830

Leu His Thr Met Leu Phe Lys Ala Leu Met Glu Gly Asn Gln Asp Asn
        835                 840                 845

Leu Asp Leu Gly Ser Gly Ala Ile Phe Tyr Arg Ala Lys Ser Leu Asp
    850                 855                 860

Gly Asn Lys Pro Thr His Pro Ala Asn Glu Ala Ile Lys Cys Arg Asn
865                 870                 875                 880
```

```
Val Ala Asn Lys Asp Lys Val Ser Leu Phe Thr Tyr Asp Ile Tyr Lys
               885                 890                 895

Asn Arg Arg Tyr Met Glu Asn Lys Phe Leu Phe His Leu Ser Ile Val
           900                 905                 910

Gln Asn Tyr Lys Ala Ala Asn Asp Ser Ala Gln Leu Asn Ser Ser Ala
        915                 920                 925

Thr Glu Tyr Ile Arg Lys Ala Asp Asp Leu His Ile Ile Gly Ile Asp
    930                 935                 940

Arg Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Val Ile Asp Met Lys Gly
945                 950                 955                 960

Asn Ile Val Glu Gln Asp Ser Leu Asn Ile Ile Arg Asn Asn Asp Leu
               965                 970                 975

Glu Thr Asp Tyr His Asp Leu Leu Asp Lys Arg Glu Lys Glu Arg Lys
           980                 985                 990

Ala Asn Arg Gln Asn Trp Glu Ala  Val Glu Gly Ile Lys  Asp Leu Lys
        995                 1000                1005

Lys Gly  Tyr Leu Ser Gln Ala  Val His Gln Ile Ala  Gln Leu Met
    1010                1015                1020

Leu Lys  Tyr Asn Ala Ile Ile  Ala Leu Glu Asp Leu  Gly Gln Met
    1025                1030                1035

Phe Val  Thr Arg Gly Gln Lys  Ile Glu Lys Ala Val  Tyr Gln Gln
    1040                1045                1050

Phe Glu  Lys Ser Leu Val Asp  Lys Leu Ser Tyr Leu  Val Asp Lys
    1055                1060                1065

Lys Arg  Pro Tyr Asn Glu Leu  Gly Gly Ile Leu Lys  Ala Tyr Gln
    1070                1075                1080

Leu Ala  Ser Ser Ile Thr Lys  Asn Asn Ser Asp Lys  Gln Asn Gly
    1085                1090                1095

Phe Leu  Phe Tyr Val Pro Ala  Trp Asn Thr Ser Lys  Ile Asp Pro
    1100                1105                1110

Val Thr  Gly Phe Thr Asp Leu  Leu Arg Pro Lys Ala  Met Thr Ile
    1115                1120                1125

Lys Glu  Ala Gln Asp Phe Phe  Gly Ala Phe Asp Asn  Ile Ser Tyr
    1130                1135                1140
```

```
Asn Asp  Lys Gly Tyr Phe Glu  Phe Glu Thr Asn Tyr  Asp Lys Phe
    1145                1150                1155

Lys Ile  Arg Met Lys Ser Ala  Gln Thr Arg Trp Thr  Ile Cys Thr
    1160                1165                1170

Phe Gly  Asn Arg Ile Lys Arg  Lys Lys Asp Lys Asn  Tyr Trp Asn
    1175                1180                1185

Tyr Glu  Glu Val Glu Leu Thr  Glu Glu Phe Lys Lys  Leu Phe Lys
    1190                1195                1200

Asp Ser  Asn Ile Asp Tyr Glu  Asn Cys Asn Leu Lys  Glu Glu Ile
    1205                1210                1215

Gln Asn  Lys Asp Asn Arg Lys  Phe Phe Asp Asp Leu  Ile Lys Leu
    1220                1225                1230

Leu Gln  Leu Thr Leu Gln Met  Arg Asn Ser Asp Asp  Lys Gly Asn
    1235                1240                1245

Asp Tyr  Ile Ile Ser Pro Val  Ala Asn Ala Glu Gly  Gln Phe Phe
    1250                1255                1260

Asp Ser  Arg Asn Gly Asp Lys  Lys Leu Pro Leu Asp  Ala Asp Ala
    1265                1270                1275

Asn Gly  Ala Tyr Asn Ile Ala  Arg Lys Gly Leu Trp  Asn Ile Arg
    1280                1285                1290

Gln Ile  Lys Gln Thr Lys Asn  Lys Asp Asp Leu Asn  Leu Ser Ile
    1295                1300                1305

Ser Ser  Thr Glu Trp Leu Asp  Phe Val Arg Glu Lys  Pro Tyr Leu
    1310                1315                1320

Lys
```

<210> 17
<211> 1484
<212> PRT
<213> Unknown

<220>
<223> Peregrinibacteria bacterium

<220>
<221> misc_feature
<222> (1073)..(1073)
<223> Xaa can be any naturally occurring amino acid

<400> 17

```
Met Ser Asn Phe Phe Lys Asn Phe Thr Asn Leu Tyr Glu Leu Ser Lys
1               5                   10                  15

Thr Leu Arg Phe Glu Leu Lys Pro Val Gly Asp Thr Leu Thr Asn Met
            20                  25                  30

Lys Asp His Leu Glu Tyr Asp Glu Lys Leu Gln Thr Phe Leu Lys Asp
        35                  40                  45

Gln Asn Ile Asp Asp Ala Tyr Gln Ala Leu Lys Pro Gln Phe Asp Glu
    50                  55                  60

Ile His Glu Glu Phe Ile Thr Asp Ser Leu Glu Ser Lys Lys Ala Lys
65                  70                  75                  80

Glu Ile Asp Phe Ser Glu Tyr Leu Asp Leu Phe Gln Glu Lys Lys Glu
                85                  90                  95

Leu Asn Asp Ser Glu Lys Lys Leu Arg Asn Lys Ile Gly Glu Thr Phe
            100                 105                 110

Asn Lys Ala Gly Glu Lys Trp Lys Lys Glu Lys Tyr Pro Gln Tyr Glu
        115                 120                 125

Trp Lys Lys Gly Ser Lys Ile Ala Asn Gly Ala Asp Ile Leu Ser Cys
    130                 135                 140

Gln Asp Met Leu Gln Phe Ile Lys Tyr Lys Asn Pro Glu Asp Glu Lys
145                 150                 155                 160

Ile Lys Asn Tyr Ile Asp Asp Thr Leu Lys Gly Phe Phe Thr Tyr Phe
                165                 170                 175

Gly Gly Phe Asn Gln Asn Arg Ala Asn Tyr Tyr Glu Thr Lys Lys Glu
            180                 185                 190

Ala Ser Thr Ala Val Ala Thr Arg Ile Val His Glu Asn Leu Pro Lys
        195                 200                 205

Phe Cys Asp Asn Val Ile Gln Phe Lys His Ile Ile Lys Arg Lys Lys
    210                 215                 220

Asp Gly Thr Val Glu Lys Thr Glu Arg Lys Thr Glu Tyr Leu Asn Ala
225                 230                 235                 240

Tyr Gln Tyr Leu Lys Asn Asn Asn Lys Ile Thr Gln Ile Lys Asp Ala
                245                 250                 255

Glu Thr Glu Lys Met Ile Glu Ser Thr Pro Ile Ala Glu Lys Ile Phe
            260                 265                 270

Asp Val Tyr Tyr Phe Ser Ser Cys Leu Ser Gln Lys Gln Ile Glu Glu
```

```
        275                 280                 285

Tyr Asn Arg Ile Ile Gly His Tyr Asn Leu Leu Ile Asn Leu Tyr Asn
    290                 295                 300

Gln Ala Lys Arg Ser Glu Gly Lys His Leu Ser Ala Asn Glu Lys Lys
305                 310                 315                 320

Tyr Lys Asp Leu Pro Lys Phe Lys Thr Leu Tyr Lys Gln Ile Gly Cys
                325                 330                 335

Gly Lys Lys Lys Asp Leu Phe Tyr Thr Ile Lys Cys Asp Thr Glu Glu
            340                 345                 350

Glu Ala Asn Lys Ser Arg Asn Glu Gly Lys Glu Ser His Ser Val Glu
        355                 360                 365

Glu Ile Ile Asn Lys Ala Gln Glu Ala Ile Asn Lys Tyr Phe Lys Ser
    370                 375                 380

Asn Asn Asp Cys Glu Asn Ile Asn Thr Val Pro Asp Phe Ile Asn Tyr
385                 390                 395                 400

Ile Leu Thr Lys Glu Asn Tyr Glu Gly Val Tyr Trp Ser Lys Ala Ala
                405                 410                 415

Met Asn Thr Ile Ser Asp Lys Tyr Phe Ala Asn Tyr His Asp Leu Gln
            420                 425                 430

Asp Arg Leu Lys Glu Ala Lys Val Phe Gln Lys Ala Asp Lys Lys Ser
        435                 440                 445

Glu Asp Asp Ile Lys Ile Pro Glu Ala Ile Glu Leu Ser Gly Leu Phe
    450                 455                 460

Gly Val Leu Asp Ser Leu Ala Asp Trp Gln Thr Thr Leu Phe Lys Ser
465                 470                 475                 480

Ser Ile Leu Ser Asn Glu Lys Leu Lys Ile Ile Thr Asp Ser Gln Thr
                485                 490                 495

Pro Ser Glu Ala Leu Leu Lys Met Ile Phe Asn Asp Ile Glu Lys Asn
            500                 505                 510

Met Glu Ser Phe Leu Lys Glu Thr Asn Asp Ile Ile Thr Leu Lys Lys
        515                 520                 525

Tyr Lys Gly Asn Lys Glu Gly Thr Glu Lys Ile Lys Gln Trp Phe Asp
    530                 535                 540

Tyr Thr Leu Ala Ile Asn Arg Met Leu Lys Tyr Phe Leu Val Lys Glu
545                 550                 555                 560
```

```
Asn Lys Ile Lys Gly Asn Ser Leu Asp Thr Asn Ile Ser Glu Ala Leu
                565                 570                 575

Lys Thr Leu Ile Tyr Ser Asp Asp Ala Glu Trp Phe Lys Trp Tyr Asp
            580                 585                 590

Ala Leu Arg Asn Tyr Leu Thr Gln Lys Pro Gln Asp Glu Ala Lys Glu
        595                 600                 605

Asn Lys Leu Lys Leu Asn Phe Asp Asn Pro Ser Leu Ala Gly Gly Trp
    610                 615                 620

Asp Val Asn Lys Glu Cys Ser Asn Phe Cys Val Ile Leu Lys Asp Lys
625                 630                 635                 640

Asn Glu Lys Lys Tyr Leu Ala Met Ile Lys Lys Gly Glu Asn Thr Leu
                645                 650                 655

Phe Gln Lys Glu Trp Thr Glu Gly Arg Gly Lys Asn Leu Thr Lys Lys
            660                 665                 670

Ser Asn Pro Leu Phe Glu Ile Asn Asn Cys Glu Ile Leu Ser Lys Met
        675                 680                 685

Glu Tyr Asp Phe Trp Ala Asp Val Ser Lys Met Ile Pro Lys Cys Ser
    690                 695                 700

Thr Gln Leu Lys Ala Val Val Asn His Phe Lys Gln Ser Asp Asn Glu
705                 710                 715                 720

Phe Ile Phe Pro Ile Gly Tyr Lys Val Thr Ser Gly Glu Lys Phe Arg
                725                 730                 735

Glu Glu Cys Lys Ile Ser Lys Gln Asp Phe Glu Leu Asn Asn Lys Val
            740                 745                 750

Phe Asn Lys Asn Glu Leu Ser Val Thr Ala Met Arg Tyr Asp Leu Ser
        755                 760                 765

Ser Thr Gln Glu Lys Gln Tyr Ile Lys Ala Phe Gln Lys Glu Tyr Trp
    770                 775                 780

Glu Leu Leu Phe Lys Gln Glu Lys Arg Asp Thr Lys Leu Thr Asn Asn
785                 790                 795                 800

Glu Ile Phe Asn Glu Trp Ile Asn Phe Cys Asn Lys Lys Tyr Ser Glu
                805                 810                 815

Leu Leu Ser Trp Glu Arg Lys Tyr Lys Asp Ala Leu Thr Asn Trp Ile
            820                 825                 830
```

```
Asn Phe Cys Lys Tyr Phe Leu Ser Lys Tyr Pro Lys Thr Thr Leu Phe
        835                 840                 845

Asn Tyr Ser Phe Lys Glu Ser Glu Asn Tyr Asn Ser Leu Asp Glu Phe
    850                 855                 860

Tyr Arg Asp Val Asp Ile Cys Ser Tyr Lys Leu Asn Ile Asn Thr Thr
865                 870                 875                 880

Ile Asn Lys Ser Ile Leu Asp Arg Leu Val Glu Glu Gly Lys Leu Tyr
                885                 890                 895

Leu Phe Glu Ile Lys Asn Gln Asp Ser Asn Asp Gly Lys Ser Ile Gly
            900                 905                 910

His Lys Asn Asn Leu His Thr Ile Tyr Trp Asn Ala Ile Phe Glu Asn
        915                 920                 925

Phe Asp Asn Arg Pro Lys Leu Asn Gly Glu Ala Glu Ile Phe Tyr Arg
    930                 935                 940

Lys Ala Ile Ser Lys Asp Lys Leu Gly Ile Val Lys Gly Lys Lys Thr
945                 950                 955                 960

Lys Asn Gly Thr Trp Ile Ile Lys Asn Tyr Arg Phe Ser Lys Glu Lys
                965                 970                 975

Phe Ile Leu His Val Pro Ile Thr Leu Asn Phe Cys Ser Asn Asn Glu
            980                 985                 990

Tyr Val Asn Asp Ile Val Asn Thr  Lys Phe Tyr Asn Phe  Ser Asn Leu
        995                 1000                 1005

His Phe  Leu Gly Ile Asp Arg  Gly Glu Lys His Leu  Ala Tyr Tyr
    1010                1015                1020

Ser Leu  Val Asn Lys Asn Gly  Glu Ile Val Asp Gln  Gly Thr Leu
    1025                1030                1035

Asn Leu  Pro Phe Thr Asp Lys  Asp Gly Asn Gln Arg  Ser Ile Lys
    1040                1045                1050

Lys Glu  Lys Tyr Phe Tyr Asn  Lys Gln Glu Asp Lys  Trp Glu Ala
    1055                1060                1065

Lys Glu  Val Asp Xaa Trp Asn  Tyr Asn Asp Leu Leu  Asp Ala Met
    1070                1075                1080

Ala Ser  Asn Arg Asp Met Ala  Arg Lys Asn Trp Gln  Arg Ile Gly
    1085                1090                1095
```

```
Thr Ile  Lys Glu Ala Lys Asn  Gly Tyr Val Ser Leu  Val Ile Arg
    1100                 1105                 1110

Lys Ile  Ala Asp Leu Ala Val  Asn Asn Glu Arg Pro  Ala Phe Ile
    1115                 1120                 1125

Val Leu  Glu Asp Leu Asn Thr  Gly Phe Lys Arg Ser  Arg Gln Lys
    1130                 1135                 1140

Ile Asp  Lys Ser Val Tyr Gln  Lys Phe Glu Leu Ala  Leu Ala Lys
    1145                 1150                 1155

Lys Leu  Asn Phe Leu Val Asp  Lys Asn Ala Lys Arg  Asp Glu Ile
    1160                 1165                 1170

Gly Ser  Pro Thr Lys Ala Leu  Gln Leu Thr Pro Pro  Val Asn Asn
    1175                 1180                 1185

Tyr Gly  Asp Ile Glu Asn Lys  Lys Gln Ala Gly Ile  Met Leu Tyr
    1190                 1195                 1200

Thr Arg  Ala Asn Tyr Thr Ser  Gln Thr Asp Pro Ala  Thr Gly Trp
    1205                 1210                 1215

Arg Lys  Thr Ile Tyr Leu Lys  Ala Gly Pro Glu Glu  Thr Thr Tyr
    1220                 1225                 1230

Lys Lys  Asp Gly Lys Ile Lys  Asn Lys Ser Val Lys  Asp Gln Ile
    1235                 1240                 1245

Ile Glu  Thr Phe Thr Asp Ile  Gly Phe Asp Gly Lys  Asp Tyr Tyr
    1250                 1255                 1260

Phe Glu  Tyr Asp Lys Gly Glu  Phe Val Asp Glu Lys  Thr Gly Glu
    1265                 1270                 1275

Ile Lys  Pro Lys Lys Trp Arg  Leu Tyr Ser Gly Glu  Asn Gly Lys
    1280                 1285                 1290

Ser Leu  Asp Arg Phe Arg Gly  Glu Arg Glu Lys Asp  Lys Tyr Glu
    1295                 1300                 1305

Trp Lys  Ile Asp Lys Ile Asp  Ile Val Lys Ile Leu  Asp Asp Leu
    1310                 1315                 1320

Phe Val  Asn Phe Asp Lys Asn  Ile Ser Leu Leu Lys  Gln Leu Lys
    1325                 1330                 1335

Glu Gly  Val Glu Leu Thr Arg  Asn Asn Glu His Gly  Thr Gly Glu
    1340                 1345                 1350

Ser Leu  Arg Phe Ala Ile Asn  Leu Ile Gln Gln Ile  Arg Asn Thr
```

```
    1355                1360                1365

Gly Asn  Asn Glu Arg Asp Asn  Asp Phe Ile Leu Ser  Pro Val Arg
    1370                1375                1380

Asp Glu  Asn Gly Lys His Phe  Asp Ser Arg Glu Tyr  Trp Asp Lys
    1385                1390                1395

Glu Thr  Lys Gly Glu Lys Ile  Ser Met Pro Ser Ser  Gly Asp Ala
    1400                1405                1410

Asn Gly  Ala Phe Asn Ile Ala  Arg Lys Gly Ile Ile  Met Asn Ala
    1415                1420                1425

His Ile  Leu Ala Asn Ser Asp  Ser Lys Asp Leu Ser  Leu Phe Val
    1430                1435                1440

Ser Asp  Glu Glu Trp Asp Leu  His Leu Asn Asn Lys  Thr Glu Trp
    1445                1450                1455

Lys Lys  Gln Leu Asn Ile Phe  Ser Ser Arg Lys Ala  Met Ala Lys
    1460                1465                1470

Arg Lys  Lys Lys Arg Pro Ala  Ala Thr Lys Lys
    1475                1480


<210>  18
<211>  1245
<212>  PRT
<213>  Porphyromonas macacae

<400>  18

Met Lys Thr Gln His Phe Phe Glu Asp Phe Thr Ser Leu Tyr Ser Leu
1               5                   10                  15

Ser Lys Thr Ile Arg Phe Glu Leu Lys Pro Ile Gly Lys Thr Leu Glu
            20                  25                  30

Asn Ile Lys Lys Asn Gly Leu Ile Arg Arg Asp Glu Gln Arg Leu Asp
        35                  40                  45

Asp Tyr Glu Lys Leu Lys Lys Val Ile Asp Glu Tyr His Glu Asp Phe
    50                  55                  60

Ile Ala Asn Ile Leu Ser Ser Phe Ser Phe Ser Glu Glu Ile Leu Gln
65                  70                  75                  80

Ser Tyr Ile Gln Asn Leu Ser Ile Ser Glu Ala Arg Ala Lys Ile Glu
                85                  90                  95

Lys Thr Met Arg Asp Thr Leu Ala Lys Ala Phe Ser Glu Asp Glu Arg
            100                 105                 110
```

```
Tyr Lys Ser Ile Phe Lys Lys Glu Leu Val Lys Lys Asp Ile Pro Val
        115                 120                 125

Trp Cys Pro Ala Tyr Lys Ser Leu Cys Lys Lys Phe Asp Asn Phe Thr
    130                 135                 140

Thr Ser Leu Val Pro Phe His Glu Asn Arg Lys Asn Leu Tyr Thr Ser
145                 150                 155                 160

Asn Glu Ile Thr Ala Ser Ile Pro Tyr Arg Ile Val His Val Asn Leu
                165                 170                 175

Pro Lys Phe Ile Gln Asn Ile Glu Ala Leu Cys Glu Leu Gln Lys Lys
            180                 185                 190

Met Gly Ala Asp Leu Tyr Leu Glu Met Met Glu Asn Leu Arg Asn Val
        195                 200                 205

Trp Pro Ser Phe Val Lys Thr Pro Asp Asp Leu Cys Asn Leu Lys Thr
    210                 215                 220

Tyr Asn His Leu Met Val Gln Ser Ser Ile Ser Glu Tyr Asn Arg Phe
225                 230                 235                 240

Val Gly Gly Tyr Ser Thr Glu Asp Gly Thr Lys His Gln Gly Ile Asn
                245                 250                 255

Glu Trp Ile Asn Ile Tyr Arg Gln Arg Asn Lys Glu Met Arg Leu Pro
            260                 265                 270

Gly Leu Val Phe Leu His Lys Gln Ile Leu Ala Lys Val Asp Ser Ser
        275                 280                 285

Ser Phe Ile Ser Asp Thr Leu Glu Asn Asp Asp Gln Val Phe Cys Val
    290                 295                 300

Leu Arg Gln Phe Arg Lys Leu Phe Trp Asn Thr Val Ser Ser Lys Glu
305                 310                 315                 320

Asp Asp Ala Ala Ser Leu Lys Asp Leu Phe Cys Gly Leu Ser Gly Tyr
                325                 330                 335

Asp Pro Glu Ala Ile Tyr Val Ser Asp Ala His Leu Ala Thr Ile Ser
            340                 345                 350

Lys Asn Ile Phe Asp Arg Trp Asn Tyr Ile Ser Asp Ala Ile Arg Arg
        355                 360                 365

Lys Thr Glu Val Leu Met Pro Arg Lys Lys Glu Ser Val Glu Arg Tyr
    370                 375                 380
```

```
Ala Glu Lys Ile Ser Lys Gln Ile Lys Lys Arg Gln Ser Tyr Ser Leu
385                 390                 395                 400

Ala Glu Leu Asp Asp Leu Leu Ala His Tyr Ser Glu Glu Ser Leu Pro
                405                 410                 415

Ala Gly Phe Ser Leu Leu Ser Tyr Phe Thr Ser Leu Gly Gly Gln Lys
            420                 425                 430

Tyr Leu Val Ser Asp Gly Glu Val Ile Leu Tyr Glu Glu Gly Ser Asn
        435                 440                 445

Ile Trp Asp Glu Val Leu Ile Ala Phe Arg Asp Leu Gln Val Ile Leu
    450                 455                 460

Asp Lys Asp Phe Thr Glu Lys Lys Leu Gly Lys Asp Glu Glu Ala Val
465                 470                 475                 480

Ser Val Ile Lys Lys Ala Leu Asp Ser Ala Leu Arg Leu Arg Lys Phe
                485                 490                 495

Phe Asp Leu Leu Ser Gly Thr Gly Ala Glu Ile Arg Arg Asp Ser Ser
            500                 505                 510

Phe Tyr Ala Leu Tyr Thr Asp Arg Met Asp Lys Leu Lys Gly Leu Leu
        515                 520                 525

Lys Met Tyr Asp Lys Val Arg Asn Tyr Leu Thr Lys Lys Pro Tyr Ser
    530                 535                 540

Ile Glu Lys Phe Lys Leu His Phe Asp Asn Pro Ser Leu Leu Ser Gly
545                 550                 555                 560

Trp Asp Lys Asn Lys Glu Leu Asn Asn Leu Ser Val Ile Phe Arg Gln
                565                 570                 575

Asn Gly Tyr Tyr Tyr Leu Gly Ile Met Thr Pro Lys Gly Lys Asn Leu
            580                 585                 590

Phe Lys Thr Leu Pro Lys Leu Gly Ala Glu Glu Met Phe Tyr Glu Lys
        595                 600                 605

Met Glu Tyr Lys Gln Ile Ala Glu Pro Met Leu Met Leu Pro Lys Val
    610                 615                 620

Phe Phe Pro Lys Lys Thr Lys Pro Ala Phe Ala Pro Asp Gln Ser Val
625                 630                 635                 640

Val Asp Ile Tyr Asn Lys Lys Thr Phe Lys Thr Gly Gln Lys Gly Phe
                645                 650                 655

Asn Lys Lys Asp Leu Tyr Arg Leu Ile Asp Phe Tyr Lys Glu Ala Leu
```

```
            660                 665                 670

Thr Val His Glu Trp Lys Leu Phe Asn Phe Ser Phe Ser Pro Thr Glu
        675                 680                 685

Gln Tyr Arg Asn Ile Gly Glu Phe Phe Asp Glu Val Arg Glu Gln Ala
    690                 695                 700

Tyr Lys Val Ser Met Val Asn Val Pro Ala Ser Tyr Ile Asp Glu Ala
705                 710                 715                 720

Val Glu Asn Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
                725                 730                 735

Ser Pro Tyr Ser Lys Gly Ile Pro Asn Leu His Thr Leu Tyr Trp Lys
            740                 745                 750

Ala Leu Phe Ser Glu Gln Asn Gln Ser Arg Val Tyr Lys Leu Cys Gly
        755                 760                 765

Gly Gly Glu Leu Phe Tyr Arg Lys Ala Ser Leu His Met Gln Asp Thr
    770                 775                 780

Thr Val His Pro Lys Gly Ile Ser Ile His Lys Lys Asn Leu Asn Lys
785                 790                 795                 800

Lys Gly Glu Thr Ser Leu Phe Asn Tyr Asp Leu Val Lys Asp Lys Arg
                805                 810                 815

Phe Thr Glu Asp Lys Phe Phe Phe His Val Pro Ile Ser Ile Asn Tyr
            820                 825                 830

Lys Asn Lys Lys Ile Thr Asn Val Asn Gln Met Val Arg Asp Tyr Ile
        835                 840                 845

Ala Gln Asn Asp Asp Leu Gln His Gly Ile Asp Arg Gly Glu Arg Asn
    850                 855                 860

Leu Leu Tyr Ile Ser Arg Ile Asp Thr Arg Gly Asn Leu Leu Glu Gln
865                 870                 875                 880

Phe Ser Leu Asn Val Ile Glu Ser Asp Lys Gly Asp Leu Arg Thr Asp
                885                 890                 895

Tyr Gln Lys Ile Leu Gly Asp Arg Glu Gln Glu Arg Leu Arg Arg Arg
            900                 905                 910

Gln Glu Trp Lys Ser Ile Glu Ser Ile Lys Asp Leu Lys Asp Gly Tyr
        915                 920                 925

Met Ser Gln Val Val His Lys Ile Cys Asn Met Val Val Glu His Lys
    930                 935                 940
```

```
Ala Ile Val Val Leu Glu Asn Leu Asn Leu Ser Phe Met Lys Gly Arg
945                 950                 955                 960


Lys Lys Val Glu Lys Ser Val Tyr Glu Lys Phe Glu Arg Met Leu Val
                965                 970                 975


Asp Lys Leu Asn Tyr Leu Val Val Asp Lys Lys Asn Leu Ser Asn Glu
            980                 985                 990


Pro Gly Gly Leu Tyr Ala Ala Tyr  Gln Leu Thr Asn Pro  Leu Phe Ser
        995                 1000                1005


Phe Glu  Glu Leu His Arg Tyr  Pro Gln Ser Gly Ile  Leu Phe Phe
    1010                1015                1020


Val Asp  Pro Trp Asn Thr Ser  Leu Thr Asp Pro Ser  Thr Gly Phe
    1025                1030                1035


Val Asn  Leu Leu Gly Arg Ile  Asn Tyr Thr Asn Val  Gly Asp Ala
    1040                1045                1050


Arg Lys  Phe Phe Asp Arg Phe  Asn Ala Ile Arg Tyr  Asp Gly Lys
    1055                1060                1065


Gly Asn  Ile Leu Phe Asp Leu  Asp Leu Ser Arg Phe  Asp Val Arg
    1070                1075                1080


Val Glu  Thr Gln Arg Lys Leu  Trp Thr Leu Thr Thr  Phe Gly Ser
    1085                1090                1095


Arg Ile  Ala Lys Ser Lys Lys  Ser Gly Lys Trp Met  Val Glu Arg
    1100                1105                1110


Ile Glu  Asn Leu Ser Leu Cys  Phe Leu Glu Leu Phe  Glu Gln Phe
    1115                1120                1125


Asn Ile  Gly Tyr Arg Val Glu  Lys Asp Leu Lys Lys  Ala Ile Leu
    1130                1135                1140


Ser Gln  Asp Arg Lys Glu Phe  Tyr Val Arg Leu Ile  Tyr Leu Phe
    1145                1150                1155


Asn Leu  Met Met Gln Ile Arg  Asn Ser Asp Gly Glu  Glu Asp Tyr
    1160                1165                1170


Ile Leu  Ser Pro Ala Leu Asn  Glu Lys Asn Leu Gln  Phe Asp Ser
    1175                1180                1185


Arg Leu  Ile Glu Ala Lys Asp  Leu Pro Val Asp Ala  Asp Ala Asn
    1190                1195                1200
```

```
Gly Ala  Tyr Asn Val Ala Arg  Lys Gly Leu Met Val  Val Gln Arg
    1205                 1210                 1215


Ile Lys  Arg Gly Asp His Glu  Ser Ile His Arg Ile  Gly Arg Ala
    1220                 1225                 1230


Gln Trp  Leu Arg Tyr Val Gln  Glu Gly Ile Val Glu
    1235                 1240                 1245
```

<210> 19
<211> 1250
<212> PRT
<213> Smithella sp.

<400> 19

```
Met Gln Thr Leu Phe Glu Asn Phe Thr Asn Gln Tyr Pro Val Ser Lys
1               5                   10                  15


Thr Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Lys Asp Phe Ile
            20                  25                  30


Glu Gln Lys Gly Leu Leu Lys Lys Asp Glu Asp Arg Ala Glu Lys Tyr
        35                  40                  45


Lys Lys Val Lys Asn Ile Ile Asp Glu Tyr His Lys Asp Phe Ile Glu
    50                  55                  60


Lys Ser Leu Asn Gly Leu Lys Leu Asp Gly Leu Glu Lys Tyr Lys Thr
65                  70                  75                  80


Leu Tyr Leu Lys Gln Glu Lys Asp Asp Lys Asp Lys Lys Ala Phe Asp
                85                  90                  95


Lys Glu Lys Glu Asn Leu Arg Lys Gln Ile Ala Asn Ala Phe Arg Asn
            100                 105                 110


Asn Glu Lys Phe Lys Thr Leu Phe Ala Lys Glu Leu Ile Lys Asn Asp
        115                 120                 125


Leu Met Ser Phe Ala Cys Glu Glu Asp Lys Lys Asn Val Lys Glu Phe
    130                 135                 140


Glu Ala Phe Thr Thr Tyr Phe Thr Gly Phe His Gln Asn Arg Ala Asn
145                 150                 155                 160


Met Tyr Val Ala Asp Glu Lys Arg Thr Ala Ile Ala Ser Arg Leu Ile
                165                 170                 175


His Glu Asn Leu Pro Lys Phe Ile Asp Asn Ile Lys Ile Phe Glu Lys
            180                 185                 190


Met Lys Lys Glu Ala Pro Glu Leu Leu Ser Pro Phe Asn Gln Thr Leu
```

```
        195                 200                 205

Lys Asp Met Lys Asp Val Ile Lys Gly Thr Thr Leu Glu Glu Ile Phe
    210                 215                 220

Ser Leu Asp Tyr Phe Asn Lys Thr Leu Thr Gln Ser Gly Ile Asp Ile
225                 230                 235                 240

Tyr Asn Ser Val Ile Gly Gly Arg Thr Pro Glu Glu Gly Lys Thr Lys
                245                 250                 255

Ile Lys Gly Leu Asn Glu Tyr Ile Asn Thr Asp Phe Asn Gln Lys Gln
            260                 265                 270

Thr Asp Lys Lys Lys Arg Gln Pro Lys Phe Lys Gln Leu Tyr Lys Gln
        275                 280                 285

Ile Leu Ser Asp Arg Gln Ser Leu Ser Phe Ile Ala Glu Ala Phe Lys
    290                 295                 300

Asn Asp Thr Glu Ile Leu Glu Ala Ile Glu Lys Phe Tyr Val Asn Glu
305                 310                 315                 320

Leu Leu His Phe Ser Asn Glu Gly Lys Ser Thr Asn Val Leu Asp Ala
                325                 330                 335

Ile Lys Asn Ala Val Ser Asn Leu Glu Ser Phe Asn Leu Thr Lys Met
            340                 345                 350

Tyr Phe Arg Ser Gly Ala Ser Leu Thr Asp Val Ser Arg Lys Val Phe
        355                 360                 365

Gly Glu Trp Ser Ile Ile Asn Arg Ala Leu Asp Asn Tyr Tyr Ala Thr
    370                 375                 380

Thr Tyr Pro Ile Lys Pro Arg Glu Lys Ser Glu Lys Tyr Glu Glu Arg
385                 390                 395                 400

Lys Glu Lys Trp Leu Lys Gln Asp Phe Asn Val Ser Leu Ile Gln Thr
                405                 410                 415

Ala Ile Asp Glu Tyr Asp Asn Glu Thr Val Lys Gly Lys Asn Ser Gly
            420                 425                 430

Lys Val Ile Ala Asp Tyr Phe Ala Lys Phe Cys Asp Asp Lys Glu Thr
        435                 440                 445

Asp Leu Ile Gln Lys Val Asn Glu Gly Tyr Ile Ala Val Lys Asp Leu
    450                 455                 460

Leu Asn Thr Pro Cys Pro Glu Asn Glu Lys Leu Gly Ser Asn Lys Asp
465                 470                 475                 480
```

```
Gln Val Lys Gln Ile Lys Ala Phe Met Asp Ser Ile Met Asp Ile Met
                485                 490                 495

His Phe Val Arg Pro Leu Ser Leu Lys Asp Thr Asp Lys Glu Lys Asp
            500                 505                 510

Glu Thr Phe Tyr Ser Leu Phe Thr Pro Leu Tyr Asp His Leu Thr Gln
        515                 520                 525

Thr Ile Ala Leu Tyr Asn Lys Val Arg Asn Tyr Leu Thr Gln Lys Pro
    530                 535                 540

Tyr Ser Thr Glu Lys Ile Lys Leu Asn Phe Glu Asn Ser Thr Leu Leu
545                 550                 555                 560

Gly Gly Trp Asp Leu Asn Lys Glu Thr Asp Asn Thr Ala Ile Ile Leu
                565                 570                 575

Arg Lys Asp Asn Leu Tyr Tyr Leu Gly Ile Met Asp Lys Arg His Asn
            580                 585                 590

Arg Ile Phe Arg Asn Val Pro Lys Ala Asp Lys Lys Asp Phe Cys Tyr
        595                 600                 605

Glu Lys Met Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro
    610                 615                 620

Lys Val Phe Phe Ser Gln Ser Arg Ile Gln Glu Phe Thr Pro Ser Ala
625                 630                 635                 640

Lys Leu Leu Glu Asn Tyr Ala Asn Glu Thr His Lys Lys Gly Asp Asn
                645                 650                 655

Phe Asn Leu Asn His Cys His Lys Leu Ile Asp Phe Phe Lys Asp Ser
            660                 665                 670

Ile Asn Lys His Glu Asp Trp Lys Asn Phe Asp Phe Arg Phe Ser Ala
        675                 680                 685

Thr Ser Thr Tyr Ala Asp Leu Ser Gly Phe Tyr His Glu Val Glu His
    690                 695                 700

Gln Gly Tyr Lys Ile Ser Phe Gln Ser Val Ala Asp Ser Phe Ile Asp
705                 710                 715                 720

Asp Leu Val Asn Glu Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys
                725                 730                 735

Asp Phe Ser Pro Phe Ser Lys Gly Lys Pro Asn Leu His Thr Leu Tyr
            740                 745                 750
```

```
Trp Lys Met Leu Phe Asp Glu Asn Asn Leu Lys Asp Val Val Tyr Lys
        755                 760                 765

Leu Asn Gly Glu Ala Glu Val Phe Tyr Arg Lys Lys Ser Ile Ala Glu
    770                 775                 780

Lys Asn Thr Thr Ile His Lys Ala Asn Glu Ser Ile Ile Asn Lys Asn
785                 790                 795                 800

Pro Asp Asn Pro Lys Ala Thr Ser Thr Phe Asn Tyr Asp Ile Val Lys
                805                 810                 815

Asp Lys Arg Tyr Thr Ile Asp Lys Phe Gln Phe His Ile Pro Ile Thr
            820                 825                 830

Met Asn Phe Lys Ala Glu Gly Ile Phe Asn Met Asn Gln Arg Val Asn
        835                 840                 845

Gln Phe Leu Lys Ala Asn Pro Asp Ile Asn Ile Ile Gly Ile Asp Arg
    850                 855                 860

Gly Glu Arg His Leu Leu Tyr Tyr Ala Leu Ile Asn Gln Lys Gly Lys
865                 870                 875                 880

Ile Leu Lys Gln Asp Thr Leu Asn Val Ile Ala Asn Glu Lys Gln Lys
                885                 890                 895

Val Asp Tyr His Asn Leu Leu Asp Lys Lys Glu Gly Asp Arg Ala Thr
            900                 905                 910

Ala Arg Gln Glu Trp Gly Val Ile Glu Thr Ile Lys Glu Leu Lys Glu
        915                 920                 925

Gly Tyr Leu Ser Gln Val Ile His Lys Leu Thr Asp Leu Met Ile Glu
    930                 935                 940

Asn Asn Ala Ile Ile Val Met Glu Asp Leu Asn Phe Gly Phe Lys Arg
945                 950                 955                 960

Gly Arg Gln Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met
                965                 970                 975

Leu Ile Asp Lys Leu Asn Tyr Leu Val Asp Lys Asn Lys Lys Ala Asn
            980                 985                 990

Glu Leu Gly Gly Leu Leu Asn Ala  Phe Gln Leu Ala Asn  Lys Phe Glu
        995                 1000                 1005

Ser Phe  Gln Lys Met Gly Lys  Gln Asn Gly Phe Ile  Phe Tyr Val
    1010                1015                1020
```

```
Pro Ala  Trp Asn Thr Ser Lys  Thr Asp Pro Ala Thr  Gly Phe Ile
    1025                1030                1035


Asp Phe  Leu Lys Pro Arg Tyr  Glu Asn Leu Asn Gln  Ala Lys Asp
    1040                1045                1050


Phe Phe  Glu Lys Phe Asp Ser  Ile Arg Leu Asn Ser  Lys Ala Asp
    1055                1060                1065


Tyr Phe  Glu Phe Ala Phe Asp  Phe Lys Asn Phe Thr  Glu Lys Ala
    1070                1075                1080


Asp Gly  Gly Arg Thr Lys Trp  Thr Val Cys Thr Thr  Asn Glu Asp
    1085                1090                1095


Arg Tyr  Gln Trp Asn Arg Ala  Leu Asn Asn Asn Arg  Gly Ser Gln
    1100                1105                1110


Glu Lys  Tyr Asp Ile Thr Ala  Glu Leu Lys Ser Leu  Phe Asp Gly
    1115                1120                1125


Lys Val  Asp Tyr Lys Ser Gly  Lys Asp Leu Lys Gln  Gln Ile Ala
    1130                1135                1140


Ser Gln  Glu Ser Ala Asp Phe  Phe Lys Ala Leu Met  Lys Asn Leu
    1145                1150                1155


Ser Ile  Thr Leu Ser Leu Arg  His Asn Asn Gly Glu  Lys Gly Asp
    1160                1165                1170


Asn Glu  Gln Asp Tyr Ile Leu  Ser Pro Val Ala Asp  Ser Lys Gly
    1175                1180                1185


Arg Phe  Phe Asp Ser Arg Lys  Ala Asp Asp Asp Met  Pro Lys Asn
    1190                1195                1200


Ala Asp  Ala Asn Gly Ala Tyr  His Ile Ala Leu Lys  Gly Leu Trp
    1205                1210                1215


Cys Leu  Glu Gln Ile Ser Lys  Thr Asp Asp Leu Lys  Lys Val Lys
    1220                1225                1230


Leu Ala  Ile Ser Asn Lys Glu  Trp Leu Glu Phe Val  Gln Thr Leu
    1235                1240                1245


Lys Gly
    1250
```

<210> 20
<211> 3987
<212> DNA
<213> Artificial

<220>
<223> Cas12a polynucleotide

<400> 20
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

<210> 21
<211> 3987
<212> DNA
<213> Artificial

<220>
<223> Cas12a polynucleotide

<400> 21
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 tccccttcta 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

aggggaggccc 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 gggagaaaat 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

<210> 22
<211> 3987
<212> DNA
<213> Artificial

<220>
<223> Cas12a polynucleotide

<400> 22
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

<210> 23
<211> 228
<212> PRT
<213> Rattus norvegicus

<400> 23

Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg
1 5 10 15

Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg
20 25 30

Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser
35 40 45

Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn
50 55 60

Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg
65 70 75 80

Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser
85 90 95

Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu Phe
100 105 110

Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg Gln
115 120 125

Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr
130 135 140

Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser Pro
145 150 155 160

Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg Leu
165 170 175

Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu
180 185 190

Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile Ala

195 200 205

Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala
210 215 220

Thr Gly Leu Lys
225

<210> 24
<211> 199
<212> PRT
<213> Homo sapiens

<400> 24

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His
1 5 10 15

Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr
20 25 30

Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met
35 40 45

Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys
50 55 60

Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro
65 70 75 80

Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile
85 90 95

Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala
100 105 110

Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg
115 120 125

Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg
130 135 140

Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His
145 150 155 160

Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp
165 170 175

Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala
180 185 190

Ile Leu Gln Asn Gln Gly Asn
195

<210> 25
<211> 621
<212> DNA
<213> Petromyzon marinus

<400> 25
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

<210> 26
<211> 160
<212> PRT
<213> Artificial

<220>
<223> synthetic polypeptide

<400> 26

Phe Glu Arg Asn Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Tyr Leu
1   5   10   15

Leu Tyr Glu Ile Lys Trp Gly Lys Ser Gly Lys Leu Trp Arg His Trp
            20   25   30

Cys Gln Asn Asn Arg Thr Gln His Ala Glu Val Tyr Phe Leu Glu Asn
        35   40   45

Ile Phe Asn Ala Arg Arg Phe Asn Pro Ser Thr His Cys Ser Ile Thr
    50   55   60

Trp Tyr Leu Ser Trp Ser Pro Cys Ala Glu Cys Ser Gln Lys Ile Val
65   70   75   80

Asp Phe Leu Lys Glu His Pro Asn Val Leu Glu Ile Tyr Val Ala Arg
                85   90   95

Leu Tyr Tyr His Glu Asp Glu Arg Asn Arg Gln Gly Leu Arg Asp Leu
            100   105   110

Val Asn Ser Gly Val Thr Ile Arg Ile Met Asp Leu Pro Asp Tyr Asn

115 120 125

Tyr Cys Trp Lys Thr Phe Val Ser Asp Gln Gly Gly Asp Glu Asp Tyr
130 135 140

Trp Pro Gly His Phe Ala Pro Trp Ile Lys Gln Tyr Ser Leu Lys Leu
145 150 155 160

<210> 27
<211> 229
<212> PRT
<213> Rattus norvegicus

<400> 27

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

```
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
    210                 215                 220


Ala Thr Gly Leu Lys
225
```

<210> 28
<211> 198
<212> PRT
<213> Homo sapiens

<400> 28

```
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1               5                   10                  15


Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
            20                  25                  30


Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
        35                  40                  45


Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
    50                  55                  60


Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65                  70                  75                  80


Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
                85                  90                  95


Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
            100                 105                 110


Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
        115                 120                 125


Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
    130                 135                 140


Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys
145                 150                 155                 160


Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
                165                 170                 175


Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
            180                 185                 190


Phe Arg Thr Leu Gly Leu
        195
```

<210> 29

<211> 197
<212> PRT
<213> Artificial

<220>
<223> hAID*

<400> 29

Met Asp Ser Leu Leu Met Asn Arg Arg Glu Phe Leu Tyr Gln Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Gly Arg Thr Phe Lys
145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190

Phe Arg Thr Cys Thr
195

<210> 30
<211> 207
<212> PRT
<213> Artificial

<220>
<223> evoCDA1

<400> 30

Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr Thr
1 5 10 15

Phe Lys Lys Gln Phe Ser Asn Asn Lys Lys Ser Val Ser His Arg Cys
20 25 30

Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys Phe
35 40 45

Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly Ile
50 55 60

His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg Asp
65 70 75 80

Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro Cys
85 90 95

Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu Arg
100 105 110

Gly Asn Gly His Thr Leu Lys Ile Trp Val Cys Lys Leu Tyr Tyr Glu
115 120 125

Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn Gly
130 135 140

Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg Lys
145 150 155 160

Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp Leu
165 170 175

Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser Ile
180 185 190

Met Phe Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205

<210> 31
<211> 228
<212> PRT
<213> Artificial

<220>
<223> EvoAPOBEC1

<400> 31

Ser Ser Lys Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg Arg
1 5 10 15

Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu Arg

```
            20                  25                  30

Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His Ser
        35                  40                  45

Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val Asn
    50                  55                  60

Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr Arg
65                  70                  75                  80

Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys Ser
                85                  90                  95

Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro Asn Val Thr Leu Phe
            100                 105                 110

Ile Tyr Ile Ala Arg Leu Tyr His Leu Ala Asn Pro Arg Asn Arg Gln
        115                 120                 125

Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met Thr
    130                 135                 140

Glu Gln Glu Ser Gly Tyr Cys Trp His Asn Phe Val Asn Tyr Ser Pro
145                 150                 155                 160

Ser Asn Glu Ser His Trp Pro Arg Tyr Pro His Leu Trp Val Arg Leu
                165                 170                 175

Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu
            180                 185                 190

Asn Ile Leu Arg Arg Lys Gln Ser Gln Leu Thr Ser Phe Thr Ile Ala
        195                 200                 205

Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp Ala
    210                 215                 220

Thr Gly Leu Lys
225


<210>  32
<211>  162
<212>  PRT
<213>  Artificial

<220>
<223>  EvoFERNY

<400>  32

Ser Phe Glu Arg Asn Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Tyr
1               5                   10                  15
```

```
Leu Leu Tyr Glu Ile Lys Trp Gly Lys Ser Gly Lys Leu Trp Arg His
            20                  25                  30

Trp Cys Gln Asn Asn Arg Thr Gln His Ala Glu Val Tyr Phe Leu Glu
        35                  40                  45

Asn Ile Phe Asn Ala Arg Arg Phe Asn Pro Ser Thr His Cys Ser Ile
    50                  55                  60

Thr Trp Tyr Leu Ser Trp Ser Pro Cys Ala Glu Cys Ser Gln Lys Ile
65                  70                  75                  80

Val Asp Phe Leu Lys Glu His Pro Asn Val Asn Leu Glu Ile Tyr Val
                85                  90                  95

Ala Arg Leu Tyr Tyr Pro Glu Asn Glu Arg Asn Arg Gln Gly Leu Arg
            100                 105                 110

Asp Leu Val Asn Ser Gly Val Thr Ile Arg Ile Met Asp Leu Pro Asp
        115                 120                 125

Tyr Asn Tyr Cys Trp Lys Thr Phe Val Ser Asp Gln Gly Gly Asp Glu
    130                 135                 140

Asp Tyr Trp Pro Gly His Phe Ala Pro Trp Ile Lys Gln Tyr Ser Leu
145                 150                 155                 160

Lys Leu
```

<210> 33
<211> 166
<212> PRT
<213> Escherichia coli

<400> 33

```
Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1               5                   10                  15

Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val
            20                  25                  30

Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile
        35                  40                  45

Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
    50                  55                  60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65                  70                  75                  80

Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85                  90                  95
```

```
Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala
            100                 105                 110

Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg
        115                 120                 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
    130                 135                 140

Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys
145                 150                 155                 160

Ala Gln Ser Ser Thr Asp
                165
```

<210> 34
<211> 166
<212> PRT
<213> Artificial

<220>
<223> adenosine deaminase

<400> 34

```
Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1               5                   10                  15

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val
            20                  25                  30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
        35                  40                  45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
    50                  55                  60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65                  70                  75                  80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85                  90                  95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
            100                 105                 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
        115                 120                 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
    130                 135                 140

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys
```

145 150 155 160

Ala Gln Ser Ser Thr Asp
165

<210> 35
<211> 166
<212> PRT
<213> Artificial

<220>
<223> adenosine deaminase

<400> 35

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1 5 10 15

Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val
20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ser Ile
35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65 70 75 80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
85 90 95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
100 105 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
130 135 140

Cys Tyr Phe Phe Arg Met Arg Arg Gln Val Phe Asn Ala Gln Lys Lys
145 150 155 160

Ala Gln Ser Ser Thr Asp
165

<210> 36
<211> 166
<212> PRT
<213> Artificial

<220>
<223> adenosine deaminase

<400> 36

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1 5 10 15

Leu Ala Lys Arg Ala Leu Asp Glu Arg Glu Val Pro Val Gly Ala Val
20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
50 55 60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65 70 75 80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
85 90 95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
100 105 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
115 120 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Asn Ala Leu Leu
130 135 140

Cys Tyr Phe Phe Arg Met Arg Arg Gln Val Phe Asn Ala Gln Lys Lys
145 150 155 160

Ala Gln Ser Ser Thr Asp
165

<210> 37
<211> 166
<212> PRT
<213> Artificial

<220>
<223> adenosine deaminase

<400> 37

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1 5 10 15

Leu Ala Lys Arg Ala Leu Asp Glu Arg Glu Val Pro Val Gly Ala Val
20 25 30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
35 40 45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln

```
    50                  55                  60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65                  70                  75                  80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85                  90                  95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
            100                 105                 110

Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg
        115                 120                 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Asn Ala Leu Leu
    130                 135                 140

Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys
145                 150                 155                 160

Ala Gln Ser Ser Thr Asp
                165


<210>  38
<211>  1763
<212>  PRT
<213>  Artificial

<220>
<223>  ABE7.10

<400>  38

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1               5                   10                  15

Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val
            20                  25                  30

Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile
        35                  40                  45

Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
    50                  55                  60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65                  70                  75                  80

Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85                  90                  95

Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala
            100                 105                 110
```

```
Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg
        115                 120                 125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
    130                 135                 140

Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys
145                 150                 155                 160

Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly
                165                 170                 175

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly
            180                 185                 190

Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr Trp
        195                 200                 205

Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu
    210                 215                 220

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu
225                 230                 235                 240

Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu
                245                 250                 255

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu
            260                 265                 270

Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala
        275                 280                 285

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg
    290                 295                 300

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr
305                 310                 315                 320

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
                325                 330                 335

Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val
            340                 345                 350

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser
        355                 360                 365

Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
    370                 375                 380

Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr
```

```
385                 390                 395                 400

Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile
                405                 410                 415

Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn
            420                 425                 430

Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe
        435                 440                 445

Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg
    450                 455                 460

Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile
465                 470                 475                 480

Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu
                485                 490                 495

Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro
            500                 505                 510

Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro
        515                 520                 525

Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala
    530                 535                 540

Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg
545                 550                 555                 560

Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val
                565                 570                 575

Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu
            580                 585                 590

Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser
        595                 600                 605

Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu
    610                 615                 620

Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser
625                 630                 635                 640

Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp
                645                 650                 655

Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn
            660                 665                 670
```

```
Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala
        675                 680                 685

Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn
    690                 695                 700

Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr
705                 710                 715                 720

Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln
                725                 730                 735

Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn
            740                 745                 750

Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr
        755                 760                 765

Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu
    770                 775                 780

Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe
785                 790                 795                 800

Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala
                805                 810                 815

Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg
            820                 825                 830

Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly
        835                 840                 845

Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser
    850                 855                 860

Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly
865                 870                 875                 880

Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn
                885                 890                 895

Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr
            900                 905                 910

Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly
        915                 920                 925

Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val
    930                 935                 940
```

```
Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys
945                 950                 955                 960


Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser
                965                 970                 975


Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu
            980                 985                 990


Leu Lys Ile Ile Lys Asp Lys Asp  Phe Leu Asp Asn Glu  Glu Asn Glu
        995                 1000                1005


Asp Ile  Leu Glu Asp Ile Val  Leu Thr Leu Thr Leu  Phe Glu Asp
    1010                1015                1020


Arg Glu  Met Ile Glu Glu Arg  Leu Lys Thr Tyr Ala  His Leu Phe
    1025                1030                1035


Asp Asp  Lys Val Met Lys Gln  Leu Lys Arg Arg Arg  Tyr Thr Gly
    1040                1045                1050


Trp Gly  Arg Leu Ser Arg Lys  Leu Ile Asn Gly Ile  Arg Asp Lys
    1055                1060                1065


Gln Ser  Gly Lys Thr Ile Leu  Asp Phe Leu Lys Ser  Asp Gly Phe
    1070                1075                1080


Ala Asn  Arg Asn Phe Met Gln  Leu Ile His Asp Asp  Ser Leu Thr
    1085                1090                1095


Phe Lys  Glu Asp Ile Gln Lys  Ala Gln Val Ser Gly  Gln Gly Asp
    1100                1105                1110


Ser Leu  His Glu His Ile Ala  Asn Leu Ala Gly Ser  Pro Ala Ile
    1115                1120                1125


Lys Lys  Gly Ile Leu Gln Thr  Val Lys Val Val Asp  Glu Leu Val
    1130                1135                1140


Lys Val  Met Gly Arg His Lys  Pro Glu Asn Ile Val  Ile Glu Met
    1145                1150                1155


Ala Arg  Glu Asn Gln Thr Thr  Gln Lys Gly Gln Lys  Asn Ser Arg
    1160                1165                1170


Glu Arg  Met Lys Arg Ile Glu  Glu Gly Ile Lys Glu  Leu Gly Ser
    1175                1180                1185


Gln Ile  Leu Lys Glu His Pro  Val Glu Asn Thr Gln  Leu Gln Asn
    1190                1195                1200
```

```
Glu Lys  Leu Tyr Leu Tyr Tyr  Leu Gln Asn Gly Arg  Asp Met Tyr
    1205                 1210                 1215

Val Asp  Gln Glu Leu Asp Ile  Asn Arg Leu Ser Asp  Tyr Asp Val
    1220                 1225                 1230

Asp His  Ile Val Pro Gln Ser  Phe Leu Lys Asp Asp  Ser Ile Asp
    1235                 1240                 1245

Asn Lys  Val Leu Thr Arg Ser  Asp Lys Asn Arg Gly  Lys Ser Asp
    1250                 1255                 1260

Asn Val  Pro Ser Glu Glu Val  Val Lys Lys Met Lys  Asn Tyr Trp
    1265                 1270                 1275

Arg Gln  Leu Leu Asn Ala Lys  Leu Ile Thr Gln Arg  Lys Phe Asp
    1280                 1285                 1290

Asn Leu  Thr Lys Ala Glu Arg  Gly Gly Leu Ser Glu  Leu Asp Lys
    1295                 1300                 1305

Ala Gly  Phe Ile Lys Arg Gln  Leu Val Glu Thr Arg  Gln Ile Thr
    1310                 1315                 1320

Lys His  Val Ala Gln Ile Leu  Asp Ser Arg Met Asn  Thr Lys Tyr
    1325                 1330                 1335

Asp Glu  Asn Asp Lys Leu Ile  Arg Glu Val Lys Val  Ile Thr Leu
    1340                 1345                 1350

Lys Ser  Lys Leu Val Ser Asp  Phe Arg Lys Asp Phe  Gln Phe Tyr
    1355                 1360                 1365

Lys Val  Arg Glu Ile Asn Asn  Tyr His His Ala His  Asp Ala Tyr
    1370                 1375                 1380

Leu Asn  Ala Val Val Gly Thr  Ala Leu Ile Lys Lys  Tyr Pro Lys
    1385                 1390                 1395

Leu Glu  Ser Glu Phe Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val
    1400                 1405                 1410

Arg Lys  Met Ile Ala Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr
    1415                 1420                 1425

Ala Lys  Tyr Phe Phe Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr
    1430                 1435                 1440

Glu Ile  Thr Leu Ala Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile
    1445                 1450                 1455

Glu Thr  Asn Gly Glu Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg
```

```
    1460                1465                1470

Asp Phe  Ala Thr Val Arg Lys  Val Leu Ser Met Pro  Gln Val Asn
    1475                1480                1485

Ile Val  Lys Lys Thr Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu
    1490                1495                1500

Ser Ile  Leu Pro Lys Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys
    1505                1510                1515

Lys Asp  Trp Asp Pro Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr
    1520                1525                1530

Val Ala  Tyr Ser Val Leu Val  Val Ala Lys Val Glu  Lys Gly Lys
    1535                1540                1545

Ser Lys  Lys Leu Lys Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile
    1550                1555                1560

Met Glu  Arg Ser Ser Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu
    1565                1570                1575

Ala Lys  Gly Tyr Lys Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu
    1580                1585                1590

Pro Lys  Tyr Ser Leu Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met
    1595                1600                1605

Leu Ala  Ser Ala Gly Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu
    1610                1615                1620

Pro Ser  Lys Tyr Val Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu
    1625                1630                1635

Lys Leu  Lys Gly Ser Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe
    1640                1645                1650

Val Glu  Gln His Lys His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile
    1655                1660                1665

Ser Glu  Phe Ser Lys Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp
    1670                1675                1680

Lys Val  Leu Ser Ala Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg
    1685                1690                1695

Glu Gln  Ala Glu Asn Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu
    1700                1705                1710

Gly Ala  Pro Ala Ala Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg
    1715                1720                1725
```

Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile
    1730   1735   1740

His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
    1745   1750   1755

Gln Leu Gly Gly Asp
    1760

<210> 39
<211> 1565
<212> PRT
<213> Artificial

<220>
<223> ABE8e

<400> 39

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1   5   10   15

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val
            20   25   30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
        35   40   45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
    50   55   60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65   70   75   80

Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85   90   95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ser Lys Arg Gly Ala
            100   105   110

Ala Gly Ser Leu Met Asn Val Leu Asn Tyr Pro Gly Met Asn His Arg
        115   120   125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
    130   135   140

Cys Asp Phe Tyr Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys Lys
145   150   155   160

Ala Gln Ser Ser Ile Asn Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly
                165   170   175

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

```
            180                 185                 190

Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile
        195                 200                 205

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val
    210                 215                 220

Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile
225                 230                 235                 240

Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala
                245                 250                 255

Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
            260                 265                 270

Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala
        275                 280                 285

Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val
    290                 295                 300

Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val
305                 310                 315                 320

Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg
                325                 330                 335

Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr
            340                 345                 350

Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu
        355                 360                 365

Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln
    370                 375                 380

Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala
385                 390                 395                 400

Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser
                405                 410                 415

Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn
            420                 425                 430

Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn
        435                 440                 445

Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser
    450                 455                 460
```

```
Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly
465                 470                 475                 480

Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala
                485                 490                 495

Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala
            500                 505                 510

Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp
        515                 520                 525

Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr
    530                 535                 540

Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile
545                 550                 555                 560

Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile
                565                 570                 575

Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg
            580                 585                 590

Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro
        595                 600                 605

His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu
    610                 615                 620

Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile
625                 630                 635                 640

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn
                645                 650                 655

Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro
            660                 665                 670

Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe
        675                 680                 685

Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val
    690                 695                 700

Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu
705                 710                 715                 720

Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe
                725                 730                 735
```

```
Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr
            740                 745                 750

Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys
        755                 760                 765

Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe
    770                 775                 780

Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp
785                 790                 795                 800

Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile
                805                 810                 815

Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg
            820                 825                 830

Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu
        835                 840                 845

Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile
    850                 855                 860

Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu
865                 870                 875                 880

Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp
                885                 890                 895

Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly
            900                 905                 910

Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro
        915                 920                 925

Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
    930                 935                 940

Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met
945                 950                 955                 960

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
                965                 970                 975

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
            980                 985                 990

Leu Lys Glu His Pro Val Glu Asn  Thr Gln Leu Gln Asn  Glu Lys Leu
        995                 1000                1005
```

```
Tyr Leu  Tyr Tyr Leu Gln Asn  Gly Arg Asp Met Tyr  Val Asp Gln
    1010                1015                1020

Glu Leu  Asp Ile Asn Arg Leu  Ser Asp Tyr Asp Val  Asp His Ile
    1025                1030                1035

Val Pro  Gln Ser Phe Leu Lys  Asp Asp Ser Ile Asp  Asn Lys Val
    1040                1045                1050

Leu Thr  Arg Ser Asp Lys Asn  Arg Gly Lys Ser Asp  Asn Val Pro
    1055                1060                1065

Ser Glu  Glu Val Val Lys Lys  Met Lys Asn Tyr Trp  Arg Gln Leu
    1070                1075                1080

Leu Asn  Ala Lys Leu Ile Thr  Gln Arg Lys Phe Asp  Asn Leu Thr
    1085                1090                1095

Lys Ala  Glu Arg Gly Gly Leu  Ser Glu Leu Asp Lys  Ala Gly Phe
    1100                1105                1110

Ile Lys  Arg Gln Leu Val Glu  Thr Arg Gln Ile Thr  Lys His Val
    1115                1120                1125

Ala Gln  Ile Leu Asp Ser Arg  Met Asn Thr Lys Tyr  Asp Glu Asn
    1130                1135                1140

Asp Lys  Leu Ile Arg Glu Val  Lys Val Ile Thr Leu  Lys Ser Lys
    1145                1150                1155

Leu Val  Ser Asp Phe Arg Lys  Asp Phe Gln Phe Tyr  Lys Val Arg
    1160                1165                1170

Glu Ile  Asn Asn Tyr His His  Ala His Asp Ala Tyr  Leu Asn Ala
    1175                1180                1185

Val Val  Gly Thr Ala Leu Ile  Lys Lys Tyr Pro Lys  Leu Glu Ser
    1190                1195                1200

Glu Phe  Val Tyr Gly Asp Tyr  Lys Val Tyr Asp Val  Arg Lys Met
    1205                1210                1215

Ile Ala  Lys Ser Glu Gln Glu  Ile Gly Lys Ala Thr  Ala Lys Tyr
    1220                1225                1230

Phe Phe  Tyr Ser Asn Ile Met  Asn Phe Phe Lys Thr  Glu Ile Thr
    1235                1240                1245

Leu Ala  Asn Gly Glu Ile Arg  Lys Arg Pro Leu Ile  Glu Thr Asn
    1250                1255                1260

Gly Glu  Thr Gly Glu Ile Val  Trp Asp Lys Gly Arg  Asp Phe Ala
```

```
    1265                1270                1275

Thr Val  Arg Lys Val Leu Ser  Met Pro Gln Val Asn  Ile Val Lys
    1280                1285                1290

Lys Thr  Glu Val Gln Thr Gly  Gly Phe Ser Lys Glu  Ser Ile Leu
    1295                1300                1305

Pro Lys  Arg Asn Ser Asp Lys  Leu Ile Ala Arg Lys  Lys Asp Trp
    1310                1315                1320

Asp Pro  Lys Lys Tyr Gly Gly  Phe Asp Ser Pro Thr  Val Ala Tyr
    1325                1330                1335

Ser Val  Leu Val Val Ala Lys  Val Glu Lys Gly Lys  Ser Lys Lys
    1340                1345                1350

Leu Lys  Ser Val Lys Glu Leu  Leu Gly Ile Thr Ile  Met Glu Arg
    1355                1360                1365

Ser Ser  Phe Glu Lys Asn Pro  Ile Asp Phe Leu Glu  Ala Lys Gly
    1370                1375                1380

Tyr Lys  Glu Val Lys Lys Asp  Leu Ile Ile Lys Leu  Pro Lys Tyr
    1385                1390                1395

Ser Leu  Phe Glu Leu Glu Asn  Gly Arg Lys Arg Met  Leu Ala Ser
    1400                1405                1410

Ala Gly  Glu Leu Gln Lys Gly  Asn Glu Leu Ala Leu  Pro Ser Lys
    1415                1420                1425

Tyr Val  Asn Phe Leu Tyr Leu  Ala Ser His Tyr Glu  Lys Leu Lys
    1430                1435                1440

Gly Ser  Pro Glu Asp Asn Glu  Gln Lys Gln Leu Phe  Val Glu Gln
    1445                1450                1455

His Lys  His Tyr Leu Asp Glu  Ile Ile Glu Gln Ile  Ser Glu Phe
    1460                1465                1470

Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn Leu Asp  Lys Val Leu
    1475                1480                1485

Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro Ile Arg  Glu Gln Ala
    1490                1495                1500

Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr Asn Leu  Gly Ala Pro
    1505                1510                1515

Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile Asp Arg  Lys Arg Tyr
    1520                1525                1530
```

Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser
    1535   1540   1545

Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly
    1550   1555   1560

Gly Asp
    1565

<210> 40
<211> 1565
<212> PRT
<213> Artificial

<220>
<223> ABE8.20m

<400> 40

Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1   5   10   15

Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val
            20   25   30

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile
        35   40   45

Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
    50   55   60

Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr
65   70   75   80

Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85   90   95

Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala
            100   105   110

Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg
        115   120   125

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
    130   135   140

Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys
145   150   155   160

Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly
                165   170   175

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly

```
            180                 185                 190

Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile
        195                 200                 205

Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val
    210                 215                 220

Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile
225                 230                 235                 240

Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala
                245                 250                 255

Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
            260                 265                 270

Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala
        275                 280                 285

Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val
    290                 295                 300

Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val
305                 310                 315                 320

Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg
                325                 330                 335

Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr
            340                 345                 350

Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu
        355                 360                 365

Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln
    370                 375                 380

Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala
385                 390                 395                 400

Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser
                405                 410                 415

Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn
            420                 425                 430

Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn
        435                 440                 445

Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser
    450                 455                 460
```

```
Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly
465                 470                 475                 480

Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala
                485                 490                 495

Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala
            500                 505                 510

Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp
        515                 520                 525

Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr
    530                 535                 540

Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile
545                 550                 555                 560

Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile
                565                 570                 575

Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg
            580                 585                 590

Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro
        595                 600                 605

His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu
    610                 615                 620

Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile
625                 630                 635                 640

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn
                645                 650                 655

Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro
            660                 665                 670

Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe
        675                 680                 685

Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val
    690                 695                 700

Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu
705                 710                 715                 720

Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe
                725                 730                 735
```

```
Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr
            740                 745                 750

Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys
        755                 760                 765

Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe
    770                 775                 780

Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp
785                 790                 795                 800

Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile
                805                 810                 815

Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg
            820                 825                 830

Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu
        835                 840                 845

Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile
    850                 855                 860

Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu
865                 870                 875                 880

Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp
                885                 890                 895

Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly
            900                 905                 910

Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro
        915                 920                 925

Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
    930                 935                 940

Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met
945                 950                 955                 960

Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
                965                 970                 975

Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
            980                 985                 990

Leu Lys Glu His Pro Val Glu Asn  Thr Gln Leu Gln Asn  Glu Lys Leu
        995                 1000                1005
```

```
Tyr Leu  Tyr Tyr Leu Gln Asn  Gly Arg Asp Met Tyr  Val Asp Gln
    1010                1015                1020

Glu Leu  Asp Ile Asn Arg Leu  Ser Asp Tyr Asp Val  Asp His Ile
    1025                1030                1035

Val Pro  Gln Ser Phe Leu Lys  Asp Asp Ser Ile Asp  Asn Lys Val
    1040                1045                1050

Leu Thr  Arg Ser Asp Lys Asn  Arg Gly Lys Ser Asp  Asn Val Pro
    1055                1060                1065

Ser Glu  Glu Val Val Lys Lys  Met Lys Asn Tyr Trp  Arg Gln Leu
    1070                1075                1080

Leu Asn  Ala Lys Leu Ile Thr  Gln Arg Lys Phe Asp  Asn Leu Thr
    1085                1090                1095

Lys Ala  Glu Arg Gly Gly Leu  Ser Glu Leu Asp Lys  Ala Gly Phe
    1100                1105                1110

Ile Lys  Arg Gln Leu Val Glu  Thr Arg Gln Ile Thr  Lys His Val
    1115                1120                1125

Ala Gln  Ile Leu Asp Ser Arg  Met Asn Thr Lys Tyr  Asp Glu Asn
    1130                1135                1140

Asp Lys  Leu Ile Arg Glu Val  Lys Val Ile Thr Leu  Lys Ser Lys
    1145                1150                1155

Leu Val  Ser Asp Phe Arg Lys  Asp Phe Gln Phe Tyr  Lys Val Arg
    1160                1165                1170

Glu Ile  Asn Asn Tyr His His  Ala His Asp Ala Tyr  Leu Asn Ala
    1175                1180                1185

Val Val  Gly Thr Ala Leu Ile  Lys Lys Tyr Pro Lys  Leu Glu Ser
    1190                1195                1200

Glu Phe  Val Tyr Gly Asp Tyr  Lys Val Tyr Asp Val  Arg Lys Met
    1205                1210                1215

Ile Ala  Lys Ser Glu Gln Glu  Ile Gly Lys Ala Thr  Ala Lys Tyr
    1220                1225                1230

Phe Phe  Tyr Ser Asn Ile Met  Asn Phe Phe Lys Thr  Glu Ile Thr
    1235                1240                1245

Leu Ala  Asn Gly Glu Ile Arg  Lys Arg Pro Leu Ile  Glu Thr Asn
    1250                1255                1260

Gly Glu  Thr Gly Glu Ile Val  Trp Asp Lys Gly Arg  Asp Phe Ala
```

```
    1265                1270                1275

Thr Val  Arg Lys Val Leu Ser  Met Pro Gln Val Asn  Ile Val Lys
    1280                1285                1290

Lys Thr  Glu Val Gln Thr Gly  Gly Phe Ser Lys Glu  Ser Ile Leu
    1295                1300                1305

Pro Lys  Arg Asn Ser Asp Lys  Leu Ile Ala Arg Lys  Lys Asp Trp
    1310                1315                1320

Asp Pro  Lys Lys Tyr Gly Gly  Phe Asp Ser Pro Thr  Val Ala Tyr
    1325                1330                1335

Ser Val  Leu Val Val Ala Lys  Val Glu Lys Gly Lys  Ser Lys Lys
    1340                1345                1350

Leu Lys  Ser Val Lys Glu Leu  Leu Gly Ile Thr Ile  Met Glu Arg
    1355                1360                1365

Ser Ser  Phe Glu Lys Asn Pro  Ile Asp Phe Leu Glu  Ala Lys Gly
    1370                1375                1380

Tyr Lys  Glu Val Lys Lys Asp  Leu Ile Ile Lys Leu  Pro Lys Tyr
    1385                1390                1395

Ser Leu  Phe Glu Leu Glu Asn  Gly Arg Lys Arg Met  Leu Ala Ser
    1400                1405                1410

Ala Gly  Glu Leu Gln Lys Gly  Asn Glu Leu Ala Leu  Pro Ser Lys
    1415                1420                1425

Tyr Val  Asn Phe Leu Tyr Leu  Ala Ser His Tyr Glu  Lys Leu Lys
    1430                1435                1440

Gly Ser  Pro Glu Asp Asn Glu  Gln Lys Gln Leu Phe  Val Glu Gln
    1445                1450                1455

His Lys  His Tyr Leu Asp Glu  Ile Ile Glu Gln Ile  Ser Glu Phe
    1460                1465                1470

Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn Leu Asp  Lys Val Leu
    1475                1480                1485

Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro Ile Arg  Glu Gln Ala
    1490                1495                1500

Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr Asn Leu  Gly Ala Pro
    1505                1510                1515

Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile Asp Arg  Lys Arg Tyr
    1520                1525                1530
```

```
Thr Ser  Thr Lys Glu Val Leu  Asp Ala Thr Leu Ile  His Gln Ser
    1535                1540                1545


Ile Thr  Gly Leu Tyr Glu Thr  Arg Ile Asp Leu Ser  Gln Leu Gly
    1550                1555                1560


Gly Asp
    1565
```

<210> 41
<211> 364
<212> PRT
<213> Artificial

<220>
<223> adenosine deaminase

<400> 41

```
Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr
1               5                   10                  15


Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val
            20                  25                  30


Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile
        35                  40                  45


Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln
    50                  55                  60


Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr
65                  70                  75                  80


Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser
                85                  90                  95


Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala
            100                 105                 110


Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg
        115                 120                 125


Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu
    130                 135                 140


Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys
145                 150                 155                 160


Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly
                165                 170                 175


Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly
```

```
            180                 185                 190

Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr Trp
        195                 200                 205

Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu
    210                 215                 220

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu
225                 230                 235                 240

Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu
                245                 250                 255

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu
            260                 265                 270

Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala
        275                 280                 285

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg
    290                 295                 300

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr
305                 310                 315                 320

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
                325                 330                 335

Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val
            340                 345                 350

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp
        355                 360


<210>  42
<211>  167
<212>  PRT
<213>  Artificial

<220>
<223>  adenosine deaminase

<400>  42

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1               5                   10                  15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
            20                  25                  30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
        35                  40                  45
```

```
Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
    50                  55                  60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu
65                  70                  75                  80

Tyr Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
        115                 120                 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    130                 135                 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp
                165
```

<210> 43
<211> 167
<212> PRT
<213> Artificial

<220>
<223> adenosine deaminase

<400> 43

```
Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1               5                   10                  15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
            20                  25                  30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
        35                  40                  45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
    50                  55                  60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65                  70                  75                  80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ser Lys Arg Gly
            100                 105                 110
```

```
Ala Ala Gly Ser Leu Met Asn Val Leu Asn Tyr Pro Gly Met Asn His
        115                 120                 125


Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    130                 135                 140


Leu Cys Asp Phe Tyr Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160


Lys Ala Gln Ser Ser Ile Asn
                165


<210>  44
<211>  83
<212>  PRT
<213>  Bacillus phage AR9

<400>  44

Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val
1               5                   10                  15


Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile
            20                  25                  30


Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu
        35                  40                  45


Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr
    50                  55                  60


Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile
65                  70                  75                  80


Lys Met Leu



<210>  45
<211>  66
<212>  DNA
<213>  Saccharomyces bayanus

<400>  45
ttcttgtcgt acttatagat cgctacgtta tttcaatttt gaaaatctga gtcctgggag       60

tgcgga                                                                 66


<210>  46
<211>  609
<212>  PRT
<213>  Homo sapiens

<400>  46

Met Ser Gly Trp Glu Ser Tyr Tyr Lys Thr Glu Gly Asp Glu Glu Ala
1               5                   10                  15
```

```
Glu Glu Glu Gln Glu Glu Asn Leu Glu Ala Ser Gly Asp Tyr Lys Tyr
            20                  25                  30

Ser Gly Arg Asp Ser Leu Ile Phe Leu Val Asp Ala Ser Lys Ala Met
        35                  40                  45

Phe Glu Ser Gln Ser Glu Asp Glu Leu Thr Pro Phe Asp Met Ser Ile
    50                  55                  60

Gln Cys Ile Gln Ser Val Tyr Ile Ser Lys Ile Ile Ser Ser Asp Arg
65                  70                  75                  80

Asp Leu Leu Ala Val Val Phe Tyr Gly Thr Glu Lys Asp Lys Asn Ser
                85                  90                  95

Val Asn Phe Lys Asn Ile Tyr Val Leu Gln Glu Leu Asp Asn Pro Gly
            100                 105                 110

Ala Lys Arg Ile Leu Glu Leu Asp Gln Phe Lys Gly Gln Gln Gly Gln
        115                 120                 125

Lys Arg Phe Gln Asp Met Met Gly His Gly Ser Asp Tyr Ser Leu Ser
    130                 135                 140

Glu Val Leu Trp Val Cys Ala Asn Leu Phe Ser Asp Val Gln Phe Lys
145                 150                 155                 160

Met Ser His Lys Arg Ile Met Leu Phe Thr Asn Glu Asp Asn Pro His
                165                 170                 175

Gly Asn Asp Ser Thr Lys Ala Ser Arg Ala Arg Thr Lys Ala Gly Asp
            180                 185                 190

Leu Arg Asp Thr Gly Ile Phe Leu Asp Leu Met His Leu Lys Lys Pro
        195                 200                 205

Gly Gly Phe Asp Ile Ser Leu Phe Tyr Arg Asp Ile Ile Ser Ile Ala
    210                 215                 220

Glu Asp Glu Asp Leu Arg Val His Phe Glu Glu Ser Ser Lys Leu Glu
225                 230                 235                 240

Asp Leu Leu Arg Lys Val Arg Ala Lys Glu Thr Arg Lys Arg Ala Leu
                245                 250                 255

Ser Arg Leu Lys Leu Lys Leu Asn Lys Asp Ile Val Ile Ser Val Gly
            260                 265                 270

Ile Tyr Asn Leu Val Gln Lys Ala Leu Lys Pro Pro Pro Ile Lys Leu
        275                 280                 285
```

```
Tyr Arg Glu Thr Asn Glu Pro Val Lys Thr Lys Thr Arg Thr Phe Asn
    290                 295                 300

Thr Ser Thr Gly Gly Leu Leu Leu Pro Ser Asp Thr Lys Arg Ser Gln
305                 310                 315                 320

Ile Tyr Gly Ser Arg Gln Ile Ile Leu Glu Lys Glu Glu Thr Glu Glu
                325                 330                 335

Leu Lys Arg Phe Asp Asp Pro Gly Leu Met Leu Met Gly Phe Lys Pro
            340                 345                 350

Leu Val Leu Leu Lys Lys His His Tyr Leu Arg Pro Ser Leu Phe Val
        355                 360                 365

Tyr Pro Glu Glu Ser Leu Val Ile Gly Ser Ser Thr Leu Phe Ser Ala
    370                 375                 380

Leu Leu Ile Lys Cys Leu Glu Lys Glu Val Ala Ala Leu Cys Arg Tyr
385                 390                 395                 400

Thr Pro Arg Arg Asn Ile Pro Pro Tyr Phe Val Ala Leu Val Pro Gln
                405                 410                 415

Glu Glu Glu Leu Asp Asp Gln Lys Ile Gln Val Thr Pro Pro Gly Phe
            420                 425                 430

Gln Leu Val Phe Leu Pro Phe Ala Asp Asp Lys Arg Lys Met Pro Phe
        435                 440                 445

Thr Glu Lys Ile Met Ala Thr Pro Glu Gln Val Gly Lys Met Lys Ala
    450                 455                 460

Ile Val Glu Lys Leu Arg Phe Thr Tyr Arg Ser Asp Ser Phe Glu Asn
465                 470                 475                 480

Pro Val Leu Gln Gln His Phe Arg Asn Leu Glu Ala Leu Ala Leu Asp
                485                 490                 495

Leu Met Glu Pro Glu Gln Ala Val Asp Leu Thr Leu Pro Lys Val Glu
            500                 505                 510

Ala Met Asn Lys Arg Leu Gly Ser Leu Val Asp Glu Phe Lys Glu Leu
        515                 520                 525

Val Tyr Pro Pro Asp Tyr Asn Pro Glu Gly Lys Val Thr Lys Arg Lys
    530                 535                 540

His Asp Asn Glu Gly Ser Gly Ser Lys Arg Pro Lys Val Glu Tyr Ser
545                 550                 555                 560

Glu Glu Glu Leu Lys Thr His Ile Ser Lys Gly Thr Leu Gly Lys Phe
```

565 570 575

Thr Val Pro Met Leu Lys Glu Ala Cys Arg Ala Tyr Gly Leu Lys Ser
580 585 590

Gly Leu Lys Lys Gln Glu Leu Leu Glu Ala Leu Thr Lys His Phe Gln
595 600 605

Asp

<210> 47
<211> 482
<212> PRT
<213> Artificial

<220>
<223> polypeptide

<400> 47

Met Val Arg Ser Gly Asn Lys Ala Ala Trp Leu Cys Met Asp Val Gly
1 5 10 15

Phe Thr Met Ser Asn Ser Ile Pro Gly Ile Glu Ser Pro Phe Glu Gln
20 25 30

Ala Lys Lys Val Ile Thr Met Phe Val Gln Arg Gln Val Phe Ala Glu
35 40 45

Asn Lys Asp Glu Ile Ala Leu Val Leu Phe Gly Thr Asp Gly Thr Asp
50 55 60

Asn Pro Leu Ser Gly Gly Asp Gln Tyr Gln Asn Ile Thr Val His Arg
65 70 75 80

His Leu Met Leu Pro Asp Phe Asp Leu Leu Glu Asp Ile Glu Ser Lys
85 90 95

Ile Gln Pro Gly Ser Gln Gln Ala Asp Phe Leu Asp Ala Leu Ile Val
100 105 110

Ser Met Asp Val Ile Gln His Glu Thr Ile Gly Lys Lys Phe Glu Lys
115 120 125

Arg His Ile Glu Ile Phe Thr Asp Leu Ser Ser Arg Phe Ser Lys Ser
130 135 140

Gln Leu Asp Ile Ile Ile His Ser Leu Lys Lys Cys Asp Ile Ser Glu
145 150 155 160

Arg His Ser Ile His Trp Pro Cys Arg Leu Thr Ile Gly Ser Asn Leu
165 170 175

```
Ser Ile Arg Ile Ala Ala Tyr Lys Ser Ile Leu Gln Glu Arg Val Lys
            180                 185                 190

Lys Thr Thr Trp Asp Ala Lys Thr Leu Lys Lys Glu Asp Ile Gln Lys
        195                 200                 205

Glu Thr Val Tyr Cys Leu Asn Asp Asp Asp Glu Thr Glu Val Leu Lys
    210                 215                 220

Glu Asp Ile Ile Gln Gly Phe Arg Tyr Gly Ser Asp Ile Val Pro Phe
225                 230                 235                 240

Ser Lys Val Asp Glu Glu Gln Met Lys Tyr Lys Ser Glu Gly Lys Cys
                245                 250                 255

Phe Ser Val Leu Gly Phe Cys Lys Ser Ser Gln Val Gln Arg Arg Phe
            260                 265                 270

Phe Met Gly Asn Gln Val Leu Lys Val Phe Ala Ala Arg Asp Asp Glu
        275                 280                 285

Ala Ala Ala Val Ala Leu Ser Ser Leu Ile His Ala Leu Asp Asp Leu
    290                 295                 300

Asp Ile Trp Ala Ile Val Arg Tyr Ala Tyr Asp Lys Arg Ala Asn Pro
305                 310                 315                 320

Gln Val Gly Val Ala Phe Pro His Ile Lys His Asn Tyr Glu Cys Leu
                325                 330                 335

Val Tyr Val Gln Leu Pro Phe Met Glu Asp Leu Arg Gln Tyr Met Phe
            340                 345                 350

Ser Ser Leu Lys Asn Ser Lys Lys Tyr Ala Pro Thr Glu Ala Gln Leu
        355                 360                 365

Asn Ala Val Asp Ala Leu Ile Asp Ser Met Ser Leu Ala Lys Lys Asp
    370                 375                 380

Glu Lys Thr Asp Thr Leu Glu Asp Leu Phe Pro Thr Thr Lys Ile Pro
385                 390                 395                 400

Asn Pro Arg Phe Gln Arg Leu Phe Gln Cys Leu Leu His Arg Ala Leu
                405                 410                 415

His Pro Arg Glu Pro Leu Pro Pro Ile Gln Gln His Ile Trp Asn Met
            420                 425                 430

Leu Asn Pro Pro Ala Glu Val Thr Thr Lys Ser Gln Ile Pro Leu Ser
        435                 440                 445

Lys Ile Lys Thr Leu Phe Pro Leu Ile Glu Ala Lys Lys Lys Asp Gln
```

450 455 460

Val Thr Ala Gln Glu Ile Phe Gln Asp Asn His Glu Asp Gly Pro Thr
465 470 475 480

Ala Lys

<210> 48
<211> 10
<212> DNA
<213> Methanobacterium thermoautotrophicum

<400> 48
aatttttgga 10

<210> 49
<211> 83
<212> PRT
<213> Methanobacterium thermoautotrophicum

<400> 49

Gly Ser Val Ile Asp Val Ser Ser Gln Arg Val Asn Val Gln Arg Pro
1 5 10 15

Leu Asp Ala Leu Gly Asn Ser Leu Asn Ser Pro Val Ile Ile Lys Leu
20 25 30

Lys Gly Asp Arg Glu Phe Arg Gly Val Leu Lys Ser Phe Asp Leu His
35 40 45

Met Asn Leu Val Leu Asn Asp Ala Glu Glu Leu Glu Asp Gly Glu Val
50 55 60

Thr Arg Arg Leu Gly Thr Val Leu Ile Arg Gly Asp Asn Ile Val Tyr
65 70 75 80

Ile Ser Pro

<210> 50
<211> 25
<212> DNA
<213> Bacteriophage MS2

<400> 50
gcgcacatga ggatcaccca tgtgc 25

<210> 51
<211> 116
<212> PRT
<213> Bacteriophage MS2

<400> 51

Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1 5 10 15

```
Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Ile Ala Glu
            20                  25                  30

Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser Val
        35                  40                  45

Arg Gln Ser Ser Ala Gln Asn Arg Lys Tyr Thr Ile Lys Val Glu Val
    50                  55                  60

Pro Lys Gly Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro
65                  70                  75                  80

Ile Phe Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln
                85                  90                  95

Gly Leu Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn
            100                 105                 110

Ser Gly Ile Tyr
        115
```

<210> 52
<211> 26
<212> DNA
<213> Bacteriophage PP7

<400> 52
ataaggagtt tatatggaaa ccctta 26

<210> 53
<211> 127
<212> PRT
<213> Bacteriophage PP7

<400> 53

```
Met Ser Lys Thr Ile Val Leu Ser Val Gly Glu Ala Thr Arg Thr Leu
1               5                   10                  15

Thr Glu Ile Gln Ser Thr Ala Asp Arg Gln Ile Phe Glu Glu Lys Val
            20                  25                  30

Gly Pro Leu Val Gly Arg Leu Arg Leu Thr Ala Ser Leu Arg Gln Asn
        35                  40                  45

Gly Ala Lys Thr Ala Tyr Arg Val Asn Leu Lys Leu Asp Gln Ala Asp
    50                  55                  60

Trp Asp Cys Ser Thr Ser Val Cys Gly Glu Leu Pro Lys Val Arg Tyr
65                  70                  75                  80

Thr Gln Val Trp Ser His Asp Val Thr Ile Val Ala Asn Ser Thr Glu
                85                  90                  95
```

Ala Ser Arg Lys Ser Leu Tyr Asp Leu Thr Lys Ser Leu Val Ala Thr
100 105 110

Ser Gln Val Glu Asp Leu Val Val Asn Leu Val Pro Leu Gly Arg
115 120 125

<210> 54
<211> 19
<212> DNA
<213> Shigella flexneri

<400> 54
ctgaatgcct gcgagcatc 19

<210> 55
<211> 62
<212> PRT
<213> Shigella phage

<400> 55

Met Lys Ser Ile Arg Cys Lys Asn Cys Asn Lys Leu Leu Phe Lys Ala
1 5 10 15

Asp Ser Phe Asp His Ile Glu Ile Arg Cys Pro Arg Cys Lys Arg His
20 25 30

Ile Ile Met Leu Asn Ala Cys Glu His Pro Thr Glu Lys His Cys Gly
35 40 45

Lys Arg Glu Lys Ile Thr His Ser Asp Glu Thr Val Arg Tyr
50 55 60

<210> 56
<211> 19
<212> DNA
<213> Artificial

<220>
<223> RNA Spacer

<220>
<221> misc_feature
<222> (1)..(19)
<223> wherein n is A, C, T or G

<400> 56
nnnnnnnnnn nnnnnnnnn 19

<210> 57
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Target strand

<220>

<221> misc_feature
<222> (4)..(22)
<223> wherein n is A, C, T or G

<400> 57
aaannnnnnn nnnnnnnnnn nn 22

<210> 58
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Non-target strand

<220>
<221> misc_feature
<222> (4)..(22)
<223> wherein n is A, C, T or G

<400> 58
tttnnnnnnn nnnnnnnnnn nn 22

<210> 59
<211> 24
<212> PRT
<213> Artificial

<220>
<223> GCN4 sequence

<400> 59

Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala Arg
1 5 10 15

Leu Lys Lys Gly Ser Gly Ser Gly
20

<210> 60
<211> 241
<212> PRT
<213> Artificial

<220>
<223> GCN4 sequence

<400> 60

Glu Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala
1 5 10 15

Arg Leu Lys Lys Gly Ser Gly Ser Gly Glu Glu Leu Leu Ser Lys Asn
20 25 30

Tyr His Leu Glu Asn Glu Val Ala Arg Leu Lys Lys Gly Ser Gly Ser
35 40 45

Gly Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala
50 55 60

```
Arg Leu Lys Lys Gly Ser Gly Ser Gly Glu Glu Leu Leu Ser Lys Asn
65                  70                  75                  80

Tyr His Leu Glu Asn Glu Val Ala Arg Leu Lys Lys Gly Ser Gly Ser
                85                  90                  95

Gly Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala
            100                 105                 110

Arg Leu Lys Lys Gly Ser Gly Ser Gly Glu Glu Leu Leu Ser Lys Asn
        115                 120                 125

Tyr His Leu Glu Asn Glu Val Ala Arg Leu Lys Lys Gly Ser Gly Ser
    130                 135                 140

Gly Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala
145                 150                 155                 160

Arg Leu Lys Lys Gly Ser Gly Ser Gly Glu Glu Leu Leu Ser Lys Asn
                165                 170                 175

Tyr His Leu Glu Asn Glu Val Ala Arg Leu Lys Lys Gly Ser Gly Ser
            180                 185                 190

Gly Glu Glu Leu Leu Ser Lys Asn Tyr His Leu Glu Asn Glu Val Ala
        195                 200                 205

Arg Leu Lys Lys Gly Ser Gly Ser Gly Glu Glu Leu Leu Ser Lys Asn
    210                 215                 220

Tyr His Leu Glu Asn Glu Val Ala Arg Leu Lys Lys Gly Ser Gly Ser
225                 230                 235                 240

Gly
```

<210> 61
<211> 277
<212> PRT
<213> Artificial

<220>
<223> ScFv antibody

<400> 61

```
Met Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu Ser Ala
1               5                   10                  15

Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr Gly Ala
            20                  25                  30

Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro Gly Lys
        35                  40                  45
```

```
Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro Gly Val
    50                  55                  60

Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr Leu Thr
65                  70                  75                  80

Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys Ala Leu
                85                  90                  95

Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val Glu Leu
            100                 105                 110

Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
        115                 120                 125

Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly Gly Gly
    130                 135                 140

Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val Ser Gly
145                 150                 155                 160

Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala Pro Gly
                165                 170                 175

Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile Thr Asp
            180                 185                 190

Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp Asn Gly
        195                 200                 205

Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp Asp Thr
    210                 215                 220

Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly Gln Gly
225                 230                 235                 240

Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
                245                 250                 255

Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
            260                 265                 270

Gly Gly Gly Gly Ser
        275
```

<210> 62
<211> 1636
<212> PRT
<213> Artificial

<220>
<223> dLbCpf1-8xGCN4

<400> 62

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1 5 10 15

Arg Lys Val Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser
20 25 30

Lys Thr Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn
35 40 45

Ile Asp Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp
50 55 60

Tyr Lys Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile
65 70 75 80

Asn Asp Val Leu His Ser Ile Lys Leu Lys Asn Leu Asn Asn Tyr Ile
85 90 95

Ser Leu Phe Arg Lys Lys Thr Arg Thr Glu Lys Glu Asn Lys Glu Leu
100 105 110

Glu Asn Leu Glu Ile Asn Leu Arg Lys Glu Ile Ala Lys Ala Phe Lys
115 120 125

Gly Asn Glu Gly Tyr Lys Ser Leu Phe Lys Lys Asp Ile Ile Glu Thr
130 135 140

Ile Leu Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn
145 150 155 160

Ser Phe Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg
165 170 175

Glu Asn Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg
180 185 190

Cys Ile Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe
195 200 205

Glu Lys Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys
210 215 220

Glu Lys Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly
225 230 235 240

Glu Phe Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn
245 250 255

Ala Ile Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly
260 265 270

```
Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu
        275                 280                 285

Pro Lys Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser
    290                 295                 300

Leu Ser Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu
305                 310                 315                 320

Val Phe Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile
                325                 330                 335

Lys Lys Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala
            340                 345                 350

Gly Ile Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp
        355                 360                 365

Ile Phe Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr
    370                 375                 380

Asp Asp Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu
385                 390                 395                 400

Asp Asp Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu
                405                 410                 415

Gln Leu Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu
            420                 425                 430

Lys Glu Ile Ile Ile Gln Lys Val Asp Glu Ile Tyr Lys Val Tyr Gly
        435                 440                 445

Ser Ser Glu Lys Leu Phe Asp Ala Asp Phe Val Leu Glu Lys Ser Leu
    450                 455                 460

Lys Lys Asn Asp Ala Val Val Ala Ile Met Lys Asp Leu Leu Asp Ser
465                 470                 475                 480

Val Lys Ser Phe Glu Asn Tyr Ile Lys Ala Phe Phe Gly Glu Gly Lys
                485                 490                 495

Glu Thr Asn Arg Asp Glu Ser Phe Tyr Gly Asp Phe Val Leu Ala Tyr
            500                 505                 510

Asp Ile Leu Leu Lys Val Asp His Ile Tyr Asp Ala Ile Arg Asn Tyr
        515                 520                 525

Val Thr Gln Lys Pro Tyr Ser Lys Asp Lys Phe Lys Leu Tyr Phe Gln
    530                 535                 540
```

```
Asn Pro Gln Phe Met Gly Gly Trp Asp Lys Asp Lys Glu Thr Asp Tyr
545                 550                 555                 560

Arg Ala Thr Ile Leu Arg Tyr Gly Ser Lys Tyr Tyr Leu Ala Ile Met
                565                 570                 575

Asp Lys Lys Tyr Ala Lys Cys Leu Gln Lys Ile Asp Lys Asp Asp Val
            580                 585                 590

Asn Gly Asn Tyr Glu Lys Ile Asn Tyr Lys Leu Leu Pro Gly Pro Asn
        595                 600                 605

Lys Met Leu Pro Lys Val Phe Phe Ser Lys Lys Trp Met Ala Tyr Tyr
    610                 615                 620

Asn Pro Ser Glu Asp Ile Gln Lys Ile Tyr Lys Asn Gly Thr Phe Lys
625                 630                 635                 640

Lys Gly Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe
                645                 650                 655

Phe Lys Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp
            660                 665                 670

Phe Asn Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr
        675                 680                 685

Arg Glu Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser
    690                 695                 700

Lys Lys Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe
705                 710                 715                 720

Gln Ile Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn
                725                 730                 735

Leu His Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly
            740                 745                 750

Gln Ile Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser
        755                 760                 765

Leu Lys Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala
    770                 775                 780

Asn Lys Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp
785                 790                 795                 800

Val Tyr Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile
                805                 810                 815
```

```
Pro Ile Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr
            820                 825                 830

Glu Val Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly
        835                 840                 845

Ile Ala Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly
    850                 855                 860

Lys Gly Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn
865                 870                 875                 880

Phe Asn Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys
                885                 890                 895

Lys Glu Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu
            900                 905                 910

Asn Ile Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys
        915                 920                 925

Ile Cys Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp
    930                 935                 940

Leu Asn Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val
945                 950                 955                 960

Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val
                965                 970                 975

Asp Lys Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr
            980                 985                 990

Gln Ile Thr Asn Lys Phe Glu Ser  Phe Lys Ser Met Ser  Thr Gln Asn
        995                 1000                1005

Gly Phe  Ile Phe Tyr Ile Pro  Ala Trp Leu Thr Ser  Lys Ile Asp
    1010                1015                1020

Pro Ser  Thr Gly Phe Val Asn  Leu Leu Lys Thr Lys  Tyr Thr Ser
    1025                1030                1035

Ile Ala  Asp Ser Lys Lys Phe  Ile Ser Ser Phe Asp  Arg Ile Met
    1040                1045                1050

Tyr Val  Pro Glu Glu Asp Leu  Phe Glu Phe Ala Leu  Asp Tyr Lys
    1055                1060                1065

Asn Phe  Ser Arg Thr Asp Ala  Asp Tyr Ile Lys Lys  Trp Lys Leu
    1070                1075                1080

Tyr Ser  Tyr Gly Asn Arg Ile  Arg Ile Phe Arg Asn  Pro Lys Lys
```

```
    1085                1090                1095

Asn Asn  Val Phe Asp Trp Glu  Glu Val Cys Leu Thr  Ser Ala Tyr
    1100                1105                1110

Lys Glu  Leu Phe Asn Lys Tyr  Gly Ile Asn Tyr Gln  Gln Gly Asp
    1115                1120                1125

Ile Arg  Ala Leu Leu Cys Glu  Gln Ser Asp Lys Ala  Phe Tyr Ser
    1130                1135                1140

Ser Phe  Met Ala Leu Met Ser  Leu Met Leu Gln Met  Arg Asn Ser
    1145                1150                1155

Ile Thr  Gly Arg Thr Asp Val  Asp Phe Leu Ile Ser  Pro Val Lys
    1160                1165                1170

Asn Ser  Asp Gly Ile Phe Tyr  Asp Ser Arg Asn Tyr  Glu Ala Gln
    1175                1180                1185

Glu Asn  Ala Ile Leu Pro Lys  Asn Ala Asp Ala Asn  Gly Ala Tyr
    1190                1195                1200

Asn Ile  Ala Arg Lys Val Leu  Trp Ala Ile Gly Gln  Phe Lys Lys
    1205                1210                1215

Ala Glu  Asp Glu Lys Leu Asp  Lys Val Lys Ile Ala  Ile Ser Asn
    1220                1225                1230

Lys Glu  Trp Leu Glu Tyr Ala  Gln Thr Ser Val Lys  His Gly Gly
    1235                1240                1245

Gly Gly  Ser Gly Gly Gly Gly  Ser Gly Gly Gly Gly  Ser Gly Pro
    1250                1255                1260

Lys Lys  Lys Arg Lys Val Ala  Ala Ala Gly Ser Glu  Glu Leu Leu
    1265                1270                1275

Ser Lys  Asn Tyr His Leu Glu  Asn Glu Val Ala Arg  Leu Lys Lys
    1280                1285                1290

Gly Ser  Gly Ser Gly Gly Ser  Gly Ser Gly Gly Ser  Gly Ser Gly
    1295                1300                1305

Ser Gly  Gly Ser Gly Ser Gly  Gly Ser Gly Ser Gly  Glu Glu Leu
    1310                1315                1320

Leu Ser  Lys Asn Tyr His Leu  Glu Asn Glu Val Ala  Arg Leu Lys
    1325                1330                1335

Lys Gly  Ser Gly Ser Gly Gly  Ser Gly Ser Gly Gly  Ser Gly Ser
    1340                1345                1350
```

```
Gly Ser  Gly Gly Ser Gly Ser  Gly Gly Ser Gly Ser  Gly Glu Glu
    1355                 1360                 1365

Leu Leu  Ser Lys Asn Tyr His  Leu Glu Asn Glu Val  Ala Arg Leu
    1370                 1375                 1380

Lys Lys  Gly Ser Gly Ser Gly  Gly Ser Gly Ser Gly  Gly Ser Gly
    1385                 1390                 1395

Ser Gly  Ser Gly Gly Ser Gly  Ser Gly Gly Ser Gly  Ser Gly Glu
    1400                 1405                 1410

Glu Leu  Leu Ser Lys Asn Tyr  His Leu Glu Asn Glu  Val Ala Arg
    1415                 1420                 1425

Leu Lys  Lys Gly Ser Gly Ser  Gly Gly Ser Gly Ser  Gly Gly Ser
    1430                 1435                 1440

Gly Ser  Gly Ser Gly Gly Ser  Gly Ser Gly Gly Ser  Gly Ser Gly
    1445                 1450                 1455

Glu Glu  Leu Leu Ser Lys Asn  Tyr His Leu Glu Asn  Glu Val Ala
    1460                 1465                 1470

Arg Leu  Lys Lys Gly Ser Gly  Ser Gly Gly Ser Gly  Ser Gly Gly
    1475                 1480                 1485

Ser Gly  Ser Gly Ser Gly Gly  Ser Gly Ser Gly Gly  Ser Gly Ser
    1490                 1495                 1500

Gly Glu  Glu Leu Leu Ser Lys  Asn Tyr His Leu Glu  Asn Glu Val
    1505                 1510                 1515

Ala Arg  Leu Lys Lys Gly Ser  Gly Ser Gly Gly Ser  Gly Ser Gly
    1520                 1525                 1530

Gly Ser  Gly Ser Gly Ser Gly  Gly Ser Gly Ser Gly  Gly Ser Gly
    1535                 1540                 1545

Ser Gly  Glu Glu Leu Leu Ser  Lys Asn Tyr His Leu  Glu Asn Glu
    1550                 1555                 1560

Val Ala  Arg Leu Lys Lys Gly  Ser Gly Ser Gly Gly  Ser Gly Ser
    1565                 1570                 1575

Gly Gly  Ser Gly Ser Gly Ser  Gly Gly Ser Gly Ser  Gly Gly Ser
    1580                 1585                 1590

Gly Ser  Gly Glu Glu Leu Leu  Ser Lys Asn Tyr His  Leu Glu Asn
    1595                 1600                 1605
```

Glu Val Ala Arg Leu Lys Lys Ser Gly Gly Ser Lys Arg Thr Ala
    1610   1615   1620

Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
    1625   1630   1635

<210> 63
<211> 610
<212> PRT
<213> Artificial

<220>
<223> scFv-APOBEC1-GB1

<400> 63

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1   5   10   15

Arg Lys Val Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu
            20   25   30

Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr
        35   40   45

Gly Ala Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro
    50   55   60

Gly Lys Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro
65   70   75   80

Gly Val Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr
                85   90   95

Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys
            100   105   110

Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val
        115   120   125

Glu Leu Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
    130   135   140

Gly Gly Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly
145   150   155   160

Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val
                165   170   175

Ser Gly Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala
            180   185   190

Pro Gly Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile
        195   200   205

```
Thr Asp Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp
    210                 215                 220

Asn Gly Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp
225                 230                 235                 240

Asp Thr Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly
                245                 250                 255

Gln Gly Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp
            260                 265                 270

Tyr Ala Gly Ser Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
        275                 280                 285

Glu Gly Gly Ser Gly Gly Ser Gly Ser Ser Ser Glu Thr Gly Pro Val
    290                 295                 300

Ala Val Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu
305                 310                 315                 320

Val Phe Phe Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr
                325                 330                 335

Glu Ile Asn Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser Gln
            340                 345                 350

Asn Thr Asn Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr Thr
        355                 360                 365

Glu Arg Tyr Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe Leu
    370                 375                 380

Ser Trp Ser Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe Leu
385                 390                 395                 400

Ser Arg Tyr Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu Tyr
                405                 410                 415

His His Ala Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile Ser
            420                 425                 430

Ser Gly Val Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys
        435                 440                 445

Trp Arg Asn Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp Pro
    450                 455                 460

Arg Tyr Pro His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys
465                 470                 475                 480
```

```
Ile Ile Leu Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln
                485                 490                 495

Pro Gln Leu Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln
            500                 505                 510

Arg Leu Pro Pro His Ile Leu Trp Ala Thr Gly Leu Lys Ser Gly Gly
        515                 520                 525

Ser Gly Gly Ser Gly Gly Ser Tyr Lys Leu Ile Leu Asn Gly Lys Thr
    530                 535                 540

Leu Lys Gly Glu Thr Thr Thr Glu Ala Val Asp Ala Ala Thr Ala Glu
545                 550                 555                 560

Lys Val Phe Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp
                565                 570                 575

Thr Tyr Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu Ser Gly Gly
            580                 585                 590

Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg
        595                 600                 605

Lys Val
    610
```

<210> 64
<211> 578
<212> PRT
<213> Artificial

<220>
<223> scFv-hAID*Delta-GB1

<400> 64

```
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                   10                  15

Arg Lys Val Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu
            20                  25                  30

Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr
        35                  40                  45

Gly Ala Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro
    50                  55                  60

Gly Lys Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro
65                  70                  75                  80

Gly Val Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr
                85                  90                  95
```

```
Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys
            100                 105                 110

Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val
        115                 120                 125

Glu Leu Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
    130                 135                 140

Gly Gly Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly
145                 150                 155                 160

Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val
                165                 170                 175

Ser Gly Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala
            180                 185                 190

Pro Gly Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile
        195                 200                 205

Thr Asp Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp
    210                 215                 220

Asn Gly Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp
225                 230                 235                 240

Asp Thr Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly
                245                 250                 255

Gln Gly Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp
            260                 265                 270

Tyr Ala Gly Ser Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
        275                 280                 285

Glu Gly Gly Ser Gly Gly Ser Gly Ser Asp Ser Leu Leu Met Asn Arg
    290                 295                 300

Arg Glu Phe Leu Tyr Gln Phe Lys Asn Val Arg Trp Ala Lys Gly Arg
305                 310                 315                 320

Arg Glu Thr Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr
                325                 330                 335

Ser Phe Ser Leu Asp Phe Gly Tyr Leu Arg Asn Lys Asn Gly Cys His
            340                 345                 350

Val Glu Leu Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro
        355                 360                 365
```

```
Gly Arg Cys Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Tyr
    370                 375                 380

Asp Cys Ala Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu
385                 390                 395                 400

Ser Leu Arg Ile Phe Thr Ala Arg Leu Tyr Phe Cys Glu Asp Arg Lys
                405                 410                 415

Ala Glu Pro Glu Gly Leu Arg Arg Leu His Arg Ala Gly Val Gln Ile
            420                 425                 430

Ala Ile Met Thr Phe Lys Asp Tyr Phe Tyr Cys Trp Asn Thr Phe Val
        435                 440                 445

Glu Asn His Gly Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu Asn
    450                 455                 460

Ser Val Arg Leu Ser Arg Gln Leu Arg Arg Ile Leu Leu Pro Leu Tyr
465                 470                 475                 480

Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Cys Thr Ser Gly Gly
                485                 490                 495

Ser Gly Gly Ser Gly Gly Ser Tyr Lys Leu Ile Leu Asn Gly Lys Thr
            500                 505                 510

Leu Lys Gly Glu Thr Thr Thr Glu Ala Val Asp Ala Ala Thr Ala Glu
        515                 520                 525

Lys Val Phe Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp
    530                 535                 540

Thr Tyr Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu Ser Gly Gly
545                 550                 555                 560

Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg
                565                 570                 575

Lys Val
```

<210> 65
<211> 580
<212> PRT
<213> Artificial

<220>
<223> scFv-hAPOBEC3A-GB1

<400> 65

```
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                   10                  15
```

```
Arg Lys Val Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu
            20                  25                  30

Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr
        35                  40                  45

Gly Ala Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro
    50                  55                  60

Gly Lys Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro
65                  70                  75                  80

Gly Val Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr
                85                  90                  95

Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys
            100                 105                 110

Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val
        115                 120                 125

Glu Leu Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
    130                 135                 140

Gly Gly Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly
145                 150                 155                 160

Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val
                165                 170                 175

Ser Gly Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala
            180                 185                 190

Pro Gly Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile
        195                 200                 205

Thr Asp Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp
    210                 215                 220

Asn Gly Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp
225                 230                 235                 240

Asp Thr Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly
                245                 250                 255

Gln Gly Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp
            260                 265                 270

Tyr Ala Gly Ser Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
        275                 280                 285
```

```
Glu Gly Gly Ser Gly Gly Ser Gly Ser Glu Ala Ser Pro Ala Ser Gly
    290                 295                 300

Pro Arg His Leu Met Asp Pro His Ile Phe Thr Ser Asn Phe Asn Asn
305                 310                 315                 320

Gly Ile Gly Arg His Lys Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu
                325                 330                 335

Asp Asn Gly Thr Ser Val Lys Met Asp Gln His Arg Gly Phe Leu His
            340                 345                 350

Asn Gln Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu
        355                 360                 365

Leu Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln
    370                 375                 380

Ile Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp
385                 390                 395                 400

Gly Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val
                405                 410                 415

Arg Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr
            420                 425                 430

Lys Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile
        435                 440                 445

Met Thr Tyr Asp Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His
    450                 455                 460

Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln
465                 470                 475                 480

Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn Ser
                485                 490                 495

Gly Gly Ser Gly Gly Ser Gly Gly Ser Tyr Lys Leu Ile Leu Asn Gly
            500                 505                 510

Lys Thr Leu Lys Gly Glu Thr Thr Thr Glu Ala Val Asp Ala Ala Thr
        515                 520                 525

Ala Glu Lys Val Phe Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly
    530                 535                 540

Glu Trp Thr Tyr Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu Ser
545                 550                 555                 560
```

```
Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys
                565                 570                 575


Lys Arg Lys Val
            580
```

<210> 66
<211> 589
<212> PRT
<213> Artificial

<220>
<223> scFv-pmCDA1-GB1

<400> 66

```
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                   10                  15


Arg Lys Val Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu
            20                  25                  30


Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr
        35                  40                  45


Gly Ala Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro
    50                  55                  60


Gly Lys Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro
65                  70                  75                  80


Gly Val Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr
                85                  90                  95


Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys
            100                 105                 110


Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val
        115                 120                 125


Glu Leu Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
    130                 135                 140


Gly Gly Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly
145                 150                 155                 160


Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val
                165                 170                 175


Ser Gly Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala
            180                 185                 190


Pro Gly Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile
        195                 200                 205
```

```
Thr Asp Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp
    210                 215                 220

Asn Gly Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp
225                 230                 235                 240

Asp Thr Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly
                245                 250                 255

Gln Gly Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp
            260                 265                 270

Tyr Ala Gly Ser Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
        275                 280                 285

Glu Gly Gly Ser Gly Gly Ser Gly Ser Thr Asp Ala Glu Tyr Val Arg
    290                 295                 300

Ile His Glu Lys Leu Asp Ile Tyr Thr Phe Lys Lys Gln Phe Phe Asn
305                 310                 315                 320

Asn Lys Lys Ser Val Ser His Arg Cys Tyr Val Leu Phe Glu Leu Lys
                325                 330                 335

Arg Arg Gly Glu Arg Arg Ala Cys Phe Trp Gly Tyr Ala Val Asn Lys
            340                 345                 350

Pro Gln Ser Gly Thr Glu Arg Gly Ile His Ala Glu Ile Phe Ser Ile
        355                 360                 365

Arg Lys Val Glu Glu Tyr Leu Arg Asp Asn Pro Gly Gln Phe Thr Ile
    370                 375                 380

Asn Trp Tyr Ser Ser Trp Ser Pro Cys Ala Asp Cys Ala Glu Lys Ile
385                 390                 395                 400

Leu Glu Trp Tyr Asn Gln Glu Leu Arg Gly Asn Gly His Thr Leu Lys
                405                 410                 415

Ile Trp Ala Cys Lys Leu Tyr Tyr Glu Lys Asn Ala Arg Asn Gln Ile
            420                 425                 430

Gly Leu Trp Asn Leu Arg Asp Asn Gly Val Gly Leu Asn Val Met Val
        435                 440                 445

Ser Glu His Tyr Gln Cys Cys Arg Lys Ile Phe Ile Gln Ser Ser His
    450                 455                 460

Asn Gln Leu Asn Glu Asn Arg Trp Leu Glu Lys Thr Leu Lys Arg Ala
465                 470                 475                 480
```

```
Glu Lys Arg Arg Ser Glu Leu Ser Ile Met Ile Gln Val Lys Ile Leu
                485                 490                 495

His Thr Thr Lys Ser Pro Ala Val Ser Gly Gly Ser Gly Gly Ser Gly
            500                 505                 510

Gly Ser Tyr Lys Leu Ile Leu Asn Gly Lys Thr Leu Lys Gly Glu Thr
        515                 520                 525

Thr Thr Glu Ala Val Asp Ala Ala Thr Ala Glu Lys Val Phe Lys Gln
    530                 535                 540

Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp Thr Tyr Asp Asp Ala
545                 550                 555                 560

Thr Lys Thr Phe Thr Val Thr Glu Ser Gly Gly Ser Lys Arg Thr Ala
                565                 570                 575

Asp Gly Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
            580                 585
```

<210> 67
<211> 763
<212> PRT
<213> Artificial

<220>
<223> scFv-hAPOBEC3B-GB1

<400> 67

```
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                   10                  15

Arg Lys Val Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu
            20                  25                  30

Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr
        35                  40                  45

Gly Ala Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro
    50                  55                  60

Gly Lys Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro
65                  70                  75                  80

Gly Val Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr
                85                  90                  95

Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys
            100                 105                 110

Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val
        115                 120                 125
```

```
Glu Leu Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
    130                 135                 140

Gly Gly Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly
145                 150                 155                 160

Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val
                165                 170                 175

Ser Gly Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala
            180                 185                 190

Pro Gly Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile
        195                 200                 205

Thr Asp Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp
    210                 215                 220

Asn Gly Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp
225                 230                 235                 240

Asp Thr Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly
                245                 250                 255

Gln Gly Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp
            260                 265                 270

Tyr Ala Gly Ser Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
        275                 280                 285

Glu Gly Gly Ser Gly Gly Ser Gly Ser Asn Pro Gln Ile Arg Asn Pro
    290                 295                 300

Met Glu Arg Met Tyr Arg Asp Thr Phe Tyr Asp Asn Phe Glu Asn Glu
305                 310                 315                 320

Pro Ile Leu Tyr Gly Arg Ser Tyr Thr Trp Leu Cys Tyr Glu Val Lys
                325                 330                 335

Ile Lys Arg Gly Arg Ser Asn Leu Leu Trp Asp Thr Gly Val Phe Arg
            340                 345                 350

Gly Gln Val Tyr Phe Lys Pro Gln Tyr His Ala Glu Met Cys Phe Leu
        355                 360                 365

Ser Trp Phe Cys Gly Asn Gln Leu Pro Ala Tyr Lys Cys Phe Gln Ile
    370                 375                 380

Thr Trp Phe Val Ser Trp Thr Pro Cys Pro Asp Cys Val Ala Lys Leu
385                 390                 395                 400
```

```
Ala Glu Phe Leu Ser Glu His Pro Asn Val Thr Leu Thr Ile Ser Ala
                405                 410                 415

Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg Asp Tyr Arg Arg Ala Leu Cys
            420                 425                 430

Arg Leu Ser Gln Ala Gly Ala Arg Val Lys Ile Met Asp Tyr Glu Glu
        435                 440                 445

Phe Ala Tyr Cys Trp Glu Asn Phe Val Tyr Asn Glu Gly Gln Gln Phe
    450                 455                 460

Met Pro Trp Tyr Lys Phe Asp Glu Asn Tyr Ala Phe Leu His Arg Thr
465                 470                 475                 480

Leu Lys Glu Ile Leu Arg Tyr Leu Met Asp Pro Asp Thr Phe Thr Phe
                485                 490                 495

Asn Phe Asn Asn Asp Pro Leu Val Leu Arg Arg Arg Gln Thr Tyr Leu
            500                 505                 510

Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Trp Val Leu Met Asp
        515                 520                 525

Gln His Met Gly Phe Leu Cys Asn Glu Ala Lys Asn Leu Leu Cys Gly
    530                 535                 540

Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro Ser
545                 550                 555                 560

Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile Ser
                565                 570                 575

Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala Phe
            580                 585                 590

Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg Ile
        595                 600                 605

Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg Asp
    610                 615                 620

Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Glu Tyr Cys
625                 630                 635                 640

Trp Asp Thr Phe Val Tyr Arg Gln Gly Cys Pro Phe Gln Pro Trp Asp
                645                 650                 655

Gly Leu Glu Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile
            660                 665                 670

Leu Gln Asn Gln Gly Asn Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
```

```
        675                 680                 685

Tyr Lys Leu Ile Leu Asn Gly Lys Thr Leu Lys Gly Glu Thr Thr Thr
    690                 695                 700

Glu Ala Val Asp Ala Ala Thr Ala Glu Lys Val Phe Lys Gln Tyr Ala
705                 710                 715                 720

Asn Asp Asn Gly Val Asp Gly Glu Trp Thr Tyr Asp Asp Ala Thr Lys
                725                 730                 735

Thr Phe Thr Val Thr Glu Ser Gly Gly Ser Lys Arg Thr Ala Asp Gly
            740                 745                 750

Ser Glu Phe Glu Pro Lys Lys Lys Arg Lys Val
        755                 760
```

<210> 68
<211> 23
<212> DNA
<213> Homo sapiens

<400> 68
accttggaga cggcgactct ctg 23

<210> 69
<211> 23
<212> DNA
<213> Homo sapiens

<400> 69
gcggatgttc caatcagtac gca 23

<210> 70
<211> 23
<212> DNA
<213> Homo sapiens

<400> 70
tgtcaccaat cctgtcccta gtg 23

<210> 71
<211> 23
<212> DNA
<213> Homo sapiens

<400> 71
tctgtcccct ccaccccaca gtg 23

<210> 72
<211> 548
<212> PRT
<213> Artificial

<220>
<223> scFv-TadA8e-GB1

<400> 72

```
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                   10                  15

Arg Lys Val Gly Pro Asp Ile Val Met Thr Gln Ser Pro Ser Ser Leu
            20                  25                  30

Ser Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ser Ser Thr
        35                  40                  45

Gly Ala Val Thr Thr Ser Asn Tyr Ala Ser Trp Val Gln Glu Lys Pro
    50                  55                  60

Gly Lys Leu Phe Lys Gly Leu Ile Gly Gly Thr Asn Asn Arg Ala Pro
65                  70                  75                  80

Gly Val Pro Ser Arg Phe Ser Gly Ser Leu Ile Gly Asp Lys Ala Thr
                85                  90                  95

Leu Thr Ile Ser Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Phe Cys
            100                 105                 110

Ala Leu Trp Tyr Ser Asn His Trp Val Phe Gly Gln Gly Thr Lys Val
        115                 120                 125

Glu Leu Lys Arg Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
    130                 135                 140

Gly Gly Ser Ser Gly Gly Gly Ser Glu Val Lys Leu Leu Glu Ser Gly
145                 150                 155                 160

Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Lys Leu Ser Cys Ala Val
                165                 170                 175

Ser Gly Phe Ser Leu Thr Asp Tyr Gly Val Asn Trp Val Arg Gln Ala
            180                 185                 190

Pro Gly Arg Gly Leu Glu Trp Ile Gly Val Ile Trp Gly Asp Gly Ile
        195                 200                 205

Thr Asp Tyr Asn Ser Ala Leu Lys Asp Arg Phe Ile Ile Ser Lys Asp
    210                 215                 220

Asn Gly Lys Asn Thr Val Tyr Leu Gln Met Ser Lys Val Arg Ser Asp
225                 230                 235                 240

Asp Thr Ala Leu Tyr Tyr Cys Val Thr Gly Leu Phe Asp Tyr Trp Gly
                245                 250                 255

Gln Gly Thr Leu Val Thr Val Ser Ser Tyr Pro Tyr Asp Val Pro Asp
            260                 265                 270

Tyr Ala Gly Ser Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
```

```
        275                 280                 285

Glu Gly Gly Ser Gly Gly Ser Gly Ser Ser Glu Val Glu Phe Ser His
    290                 295                 300

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp
305                 310                 315                 320

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val
                325                 330                 335

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala
            340                 345                 350

His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn
        355                 360                 365

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val
    370                 375                 380

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe
385                 390                 395                 400

Gly Val Arg Asn Ser Lys Arg Gly Ala Ala Gly Ser Leu Met Asn Val
                405                 410                 415

Leu Asn Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile
            420                 425                 430

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Asp Phe Tyr Arg Met Pro
        435                 440                 445

Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Ile Asn Ser
    450                 455                 460

Gly Gly Ser Gly Gly Ser Gly Gly Ser Tyr Lys Leu Ile Leu Asn Gly
465                 470                 475                 480

Lys Thr Leu Lys Gly Glu Thr Thr Thr Glu Ala Val Asp Ala Ala Thr
                485                 490                 495

Ala Glu Lys Val Phe Lys Gln Tyr Ala Asn Asp Asn Gly Val Asp Gly
            500                 505                 510

Glu Trp Thr Tyr Asp Asp Ala Thr Lys Thr Phe Thr Val Thr Glu Ser
        515                 520                 525

Gly Gly Ser Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Pro Lys Lys
    530                 535                 540

Lys Arg Lys Val
545
```

<210> 73
<211> 23
<212> DNA
<213> Homo sapiens

<400> 73
tatgagttac aacgaacacc tca 23

<210> 74
<211> 23
<212> DNA
<213> Homo sapiens

<400> 74
cagctattca ggctggcccg ccc 23

<210> 75
<211> 23
<212> DNA
<213> Homo sapiens

<400> 75
ccttcagcta aaataaagga gga 23

<210> 76
<211> 23
<212> DNA
<213> Homo sapiens

<400> 76
gctgaaggga aataaaagga aaa 23

<210> 77
<211> 23
<212> DNA
<213> Homo sapiens

<400> 77
gctcagcagg cacctgcctc agc 23