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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 April 2022 (21.04.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/081890 A1

(51) International Patent Classification:
C12N 9/78 (2006.01) *A61P 3/00* (2006.01)
C12N 15/11 (2006.01) *A61K 48/00* (2006.01)
C12N 15/62 (2006.01)

(21) International Application Number:
PCT/US2021/055057

(22) International Filing Date:
14 October 2021 (14.10.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/091,891 14 October 2020 (14.10.2020) US
63/248,081 24 September 2021 (24.09.2021) US

(71) Applicant: **BEAM THERAPEUTICS INC.** [US/US]; 26 Landsdowne Street, 2nd Floor, Cambridge, Massachusetts 02139 (US).

(72) Inventors: **CAFFERTY, Brian**; c/o Beam Therapeutics Inc., 26 Landsdowne Street, 2nd Floor, Cambridge, Massachusetts 02139 (US). **BOHNUUD, Tanggis**; c/o Beam Therapeutics Inc., 26 Landsdowne Street, 2nd Floor, Cambridge, Massachusetts 02139 (US). **CHENG, Lo-I**; c/o Beam Therapeutics Inc., 26 Landsdowne Street, 2nd Floor, Cambridge, Massachusetts 02139 (US). **PACKER, Michael**; c/o Beam Therapeutics Inc., 26 Landsdowne Street, 2nd Floor, Cambridge, Massachusetts 02139 (US). **ARATYN-SCHAUS, Yvonne**; c/o Beam Therapeutics Inc., 26 Landsdowne Street, 2nd Floor, Cambridge, Massachusetts 02139 (US).

(74) Agent: **HUNTER-ENSOR, Melissa**; c/o Greenberg Traurig, LLP, One International Place, Boston, Massachusetts 02110 (US).

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

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

Declarations under Rule 4.17:
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** COMPOSITIONS AND METHODS FOR TREATING GLYCOGEN STORAGE DISEASE TYPE 1A

(57) **Abstract:** Described and provided are adenosine base editors and compositions comprising adenosine base editors that have increased efficiency. Also described and provided are methods of using base editors comprising adenosine deaminase variants for altering mutations associated with Glycogen Storage Disease Type 1a (GSD1a).

WO 2022/081890 A1

COMPOSITIONS AND METHODS FOR TREATING GLYCOGEN STORAGE DISEASE TYPE 1A

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to and benefit of Provisional Patent Application Nos. 63/091,891, filed on October 14, 2020 and 63/248,081, filed on September 24, 2021, the contents of all of which are hereby incorporated by reference in their entireties.

SEQUENCE LISTING

This application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. The ASCII copy, created on October 14, 2021, is named 180802-047002PCT_SL and is 2,088,767 bytes in size.

BACKGROUND OF THE DISCLOSURE

For most known genetic diseases, correction of a point mutation in the target locus, rather than stochastic disruption of the gene, is needed to study or address the underlying cause of the disease. Current genome editing technologies utilizing the clustered regularly interspaced short palindromic repeat (CRISPR) system introduce double-stranded DNA breaks at a target locus as the first step to gene correction. In response to double-stranded DNA breaks, cellular DNA repair processes mostly result in random insertions or deletions (indels) at the site of DNA cleavage through non-homologous end joining. Although most genetic diseases arise from point mutations, current approaches to point mutation correction are inefficient and typically induce an abundance of random insertions and deletions (indels) at the target locus resulting from the cellular response to dsDNA breaks. Therefore, there is a need for an improved form of genome editing that is more efficient and with far fewer undesired products such as stochastic insertions or deletions (indels) or translocations.

Glycogen Storage Disease Type 1 (also known as GSD1 or Von Gierke Disease) is an inherited disorder that results in a deficiency in glycogenolysis and gluconeogenesis, with accumulation of glycogen and lipids in tissues, causing life-threatening hypoglycemia and lactic acidosis and leading to potential CNS damage and long-term liver and renal complications, such as steatosis, hepatic adenomas and hepatocellular carcinomas.

There are two types of GSD1, Type 1a (GSD1a or GSD-Ia) and Type 1b (GSD1b), which are caused by different genetic mutations. GSD1a is caused by a mutation in the glucose-6-phosphatase (G6PC) gene and affects about 80% of patients with GSD1. About one in 100,000 newborns in the US have GSD1a with about 22% of patients carrying the recessive mutation Q347* and 37% of patients carrying the recessive mutation R83C.

There are no drug therapies approved for GSD1a. Although liver transplants are curative, there are no approved therapies and the current treatment regimen involves nearly continuous cornstarch feeding. If chronically untreated, patients develop severe lactic acidosis, can progress to renal failure, and die in infancy or childhood. GSD1a is an area of significant unmet medical need. Therefore, there is a need for novel compositions and methods for treating patients with GSD1a.

SUMMARY OF THE DISCLOSURE AND EMBODIMENTS

Featured, provided and described herein are compositions and methods for the precise correction of pathogenic amino acids using a programmable nucleobase editor. In particular, the compositions and methods disclosed and described herein are useful for the treatment of Glycogen Storage Disease Type 1a (GSD1a). Thus, compositions and methods are provided for treating GSD1a using an adenosine (A) base editor (ABE) to precisely correct a single nucleotide polymorphism in the endogenous *G6PC* gene to correct a deleterious mutation (*e.g.*, Q347X, R83C).

In an aspect, an adenosine deaminase variant including a glycine (G) at amino acid position 82, a threonine (T) or an aspartic acid (D) at amino acid position 147, a serine (S) at amino acid position 154, and one or more of a histidine (H) at amino acid position 36, a tyrosine at amino acid position 76, a tyrosine at amino acid position 149, a lysine (K) at amino acid position 157, and an asparagine (N) at amino acid position 167 of the following amino acid sequence is provided, wherein the adenosine deaminase has at least about 85% identity to said amino acid sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or

corresponding alterations in another adenosine deaminase.

Another aspect provides an adenosine deaminase variant, including any of the following combinations of alterations: a) I76Y + V82G + Y147T + Q154S; b) L36H + V82G

+ Y147T + Q154S + N157K; c) V82G + Y147D + F149Y + Q154S + D167N; d) L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; e) L36H + I76Y + V82G + Y147T + Q154S + N157K; f) I76Y + V82G + Y147D + F149Y + Q154S + D167N; g) Y147D + F149Y + D167N; h) L36H; I76Y; V82G; Q154S; and N157K; i) I76Y; V82G; Q154S; or j) L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N with reference to SEQ ID NO: 1:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1) , or corresponding combinations of alterations in another adenosine deaminase.

In an embodiment of the above-delineated adenosine deaminase variants, the adenosine deaminase variant includes the following combination of alterations I76Y + V82G + Y147D + F149Y + Q154S + D167N of SEQ ID NO: 1, or corresponding alterations in another adenosine deaminase. In another embodiment, the adenosine deaminase has at least about 90% identity to SEQ ID NO: 1. In another embodiment, the adenosine deaminase has at least about 95% identity to SEQ ID NO: 1. In another embodiment, the adenosine deaminase comprises or consists essentially of SEQ ID NO: 1.

In another aspect, a fusion protein or complex including a polynucleotide programmable DNA binding domain and at least one adenosine deaminase variant domain is provided, wherein the adenosine deaminase variant domain comprises a glycine (G) at amino acid position 82, a threonine (T) or an aspartic acid (D) at amino acid position 147, a serine (S) at amino acid position 154, and one or more of a histidine (H) at amino acid position 36, a tyrosine at amino acid position 76, a tyrosine at amino acid position 149, a lysine (K) at amino acid position 157, and an asparagine (N) at amino acid position 167 of the following amino acid sequence, wherein the adenosine deaminase has at least about 85% identity to said amino acid sequence

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase. In an embodiment of the fusion protein or complex, the adenosine deaminase variant domain has at least about 90% identity to SEQ ID NO: 1. In another embodiment of the fusion protein or complex, the adenosine deaminase variant domain at least about 95% identity to SEQ ID NO: 1. In another

embodiment of the fusion protein or complex, the adenosine deaminase variant domain comprises or consists essentially of SEQ ID NO: 1.

Yet another aspect provides a fusion protein or complex including a polynucleotide programmable DNA binding domain and at least one adenosine deaminase variant domain, wherein the adenosine deaminase variant domain comprises any of the following combinations of alterations: a) I76Y + V82G + Y147T + Q154S; b) L36H + V82G + Y147T + Q154S + N157K; c) V82G + Y147D + F149Y + Q154S + D167N; d) L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; e) L36H + I76Y + V82G + Y147T + Q154S + N157K; f) I76Y + V82G + Y147D + F149Y + Q154S + D167N; g) Y147D + F149Y + D167N; h) L36H; I76Y; V82G; Q154S; and N157K; i) I76Y; V82G; Q154S; or j) L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N with reference to SEQ ID NO: 1:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding combinations of alterations in another adenosine deaminase.

In some embodiments of the above-delineated fusion protein or complex, the adenosine deaminase variant includes the following combination of alterations I76Y + V82G + Y147D + F149Y + Q154S + D167N of SEQ ID NO: 1, or corresponding alterations in another adenosine deaminase. In some embodiments, the fusion protein or complex includes one adenosine deaminase variant domain. In some embodiments, the fusion protein or complex includes a wild-type adenosine deaminase domain and an adenosine deaminase variant domain. In some embodiments, the fusion protein or complex includes a TadA*7.10 adenosine deaminase domain and an adenosine deaminase variant domain. In some embodiments, the polynucleotide programmable DNA binding domain is a Cas9 domain. In some embodiments, the Cas9 domain comprises a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9. In some embodiments, the polynucleotide programmable DNA binding domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus 1* Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof. In some embodiments, the polynucleotide programmable DNA binding domain comprises a modified SaCas9 having an altered protospacer-adjacent motif (PAM) specificity. In some embodiments, the SaCas9 has protospacer-adjacent motif (PAM) specificity for the nucleic acid sequence 5'-NNGRRT-3'. In some embodiments, the SaCas9

has specificity for the nucleic acid sequence 5'-GAGAAT-3'. In some embodiments, the SaCas9 is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 is a nickase comprising an amino acid substitution N579A or a corresponding amino acid substitution thereof. In some embodiments, the SaCas9 is a *Streptococcus pyogenes* Cas9 (SpCas9) or a variant thereof. In some embodiments of the above-delineated fusion protein or complex, the adenosine deaminase variant is capable of deaminating adenine in deoxyribonucleic acid (DNA). In some embodiments, the fusion protein or complex further includes a linker between the polynucleotide programmable DNA binding domain and the adenosine deaminase variant domain. In some embodiments, the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 359). In some embodiments, the fusion protein or complex includes one or more nuclear localization signal. In some embodiments, the nuclear localization signal is a bipartite nuclear localization signal. In some embodiments, the polynucleotide programmable DNA binding domain non-covalently associates with the deaminase.

Another aspect provides a base editor system including any of the fusion proteins or complexes as provided herein and one or more guide polynucleotides. In some embodiments, the one or more guide polynucleotides target the fusion protein to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) associated with a genetic disease. In some embodiments, the genetic disease is Glycogen Storage Disease Type 1a (GSD1a). In some embodiments, the guide polynucleotide comprises ribonucleic acid (RNA), or deoxyribonucleic acid (DNA). In some embodiments, the guide polynucleotide comprises a nucleic acid sequence: 5'-CAGUAUGGACACUGUCCAAA-3' (SEQ ID NO: 370). In some embodiments, the guide comprises or consists of one of the following nucleic acid sequences: CACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAA AACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 409), or CCACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUA AAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 410). In some embodiments, the guide polynucleotide comprises one or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises two,

three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide. In some embodiments, the guide polynucleotide comprises four modified nucleosides at the 5' end and four modified nucleosides at the 3' end of the guide. In some embodiments, the modified nucleoside comprises a 2' O-methyl or a phosphorothioate. In some embodiments, the guide polynucleotide comprises or consists essentially of one of the following sequences: mCsmAsmCsCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUC UACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 409) or mCsmCsmAsCCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAU CUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 410), wherein "m" denotes a 2' O-methyl and the "s" denotes a phosphorothioate. In some embodiments, the guide polynucleotide comprises a nucleic acid sequence: 5'-CAGUAUGGACACUGUCCAAA-3' (SEQ ID NO: 370). In some embodiments, the guide polynucleotide comprises a nucleic acid sequence, from 5' to 3', as follows: CCACCAGUAUGGACACUGUC (SEQ ID NO: 371); CACCAGUAUGGACACUGUCC (SEQ ID NO: 372); ACCAGUAUGGACACUGUCCA (SEQ ID NO: 373); CCAGUAUGGACACUGUCCAA (SEQ ID NO: 374); CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370); AGUAUGGACACUGUCCAAAG (SEQ ID NO: 375); GUAUGGACACUGUCCAAAGA (SEQ ID NO: 376); or UAUGGACACUGUCCAAAGAG (SEQ ID NO: 377). In some embodiments, the adenosine deaminase variant domain is internal to the Cas protein.

Another aspect provides a polynucleotide encoding any of the adenosine deaminase variants as provided herein, any of the fusion proteins or complexes as provided herein, or any of the base editor systems as provided herein. In an embodiment, the polynucleotide comprises one or more modified nucleosides or nucleotides. In an embodiment, the polynucleotide is DNA or RNA. In embodiments, the polynucleotide comprises a modification selected from the group consisting of 2'-*O*-methyl (2'-OMe), phosphorothioate (PS), 2'-*O*-methyl thioPACE (MSP), 2'-*O*-methyl-PACE (MP), 2'-fluoro RNA (2'-F-RNA), and constrained ethyl (S-cEt).

Another aspect provides a cell including any of the polynucleotides as provided herein.

Yet another aspect provides a cell including any of the adenosine deaminase variants as provided herein, any of the fusion proteins or complexes as provided herein and one or

more guide polynucleotides, any of the base editor systems as provided herein, or any of the polynucleotides as provided herein.

In some embodiments of the above-delineated aspects, the cell is a hepatocyte, a hepatocyte precursor, or an iPSc-derived hepatocyte. In some embodiments, the cell expresses a G6PC polypeptide. In some embodiments, the cell is from a subject having Glycogen Storage Disease Type 1a (GSD1a). In some embodiments, the cell is a mammalian cell *in vivo*, *ex vivo,* or *in vitro*. In some embodiments, the cell is a human cell. In some embodiments, the fusion protein and the one or more guide polynucleotides form a complex in the cell.

In another aspect, a method of treating a genetic disease in a subject in need thereof is provided, in which the method involves administering to a cell of the subject any of the base editor systems as provided herein or a polynucleotide encoding the base editor system.

In another aspect, a method of treating a genetic disease in a subject in need thereof is provided, in which the method involves administering to the subject any of the cells as provided herein.

In some embodiments of the above-delineated methods, after treatment the cell expresses a G6PC polypeptide capable of catalyzing the hydrolysis of D-glucose 6-phosphate to D-glucose and orthophosphate. In some embodiments, the cell is autologous, allogeneic, or xenogeneic to the subject. In some embodiments, the genetic disease is Glycogen Storage Disease Type 1a (GSD1a) and/or symptoms thereof. In an embodiment of the above-denoted treatment methods, the subject sustains at least a 24 hour fasting period after treatment.

In another aspect, a method for correcting a single nucleotide polymorphism (SNP) in a polynucleotide is provided, in which the method involves contacting a target nucleotide sequence, at least a portion of which is located in the polynucleotide or its reverse complement, with any of the base editor systems as provided herein; and editing the SNP by deaminating the SNP or its complement nucleobase upon targeting of the base editor to the target nucleotide sequence, wherein deaminating the SNP or its complement nucleobase corrects the SNP. In some embodiments, the SNP is associated with Glycogen Storage Disease Type 1a (GSD1a). In some embodiments, the SNP is in the *G6PC* gene.

In another aspect, a method of editing a glucose-6-phosphatase (G6PC) polynucleotide comprising a single nucleotide polymorphism (SNP) associated with Glycogen Storage Disease Type 1a (GSD1a) is provided, in which the method includes contacting the G6PC polynucleotide with any of the fusion proteins or complexes as provided

herein in a complex with one or more guide polynucleotides, and wherein one or more of said guide polynucleotides target said base editor to effect an A•T to G•C alteration of the SNP associated with GSD1a.

In some embodiments of the above-delineated methods, the contacting is in a cell, a eukaryotic cell, a mammalian cell, or human cell. In some embodiments, the SNP changes a glutamine (Q) to a non-glutamine (X) amino acid or changes an arginine (R) to a non-arginine (X) in a G6PC polypeptide. In some embodiments, the SNP results in expression of an G6PC polypeptide having a non-glutamine (X) amino acid at position 347 or a non-arginine (X) amino acid at position 83. In some embodiments, the base editor correction replaces the non-glutamine amino acid (X) at position 347 with a glutamine or the non-arginine amino acid (X) at position 83 with an arginine (R). In some embodiments, the SNP results in expression of a G6PC polypeptide that prematurely terminates at amino acid position 347 or at a cysteine at position 83. In some embodiments, the SNP encodes one or more of Q347X and/or R83C. In some embodiments, the SNP encodes R83C. In some embodiments, the editing results in less than than 0.5% indel formation. In some embodiments, the editing rescues G6PC catalytic activity. In some embodiments, the guide polynucleotide comprises a nucleic acid sequence: 5'-CAGUAUGGACACUGUCCAAA-3' (SEQ ID NO: 370). In some embodiments, the guide polynucleotide comprises a nucleic acid sequence, from 5' to 3', as follows: CCACCAGUAUGGACACUGUC (SEQ ID NO: 371); CACCAGUAUGGACACUGUCC (SEQ ID NO: 372); ACCAGUAUGGACACUGUCCA (SEQ ID NO: 373); CCAGUAUGGACACUGUCCAA (SEQ ID NO: 374); CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370); AGUAUGGACACUGUCCAAAG (SEQ ID NO: 375); GUAUGGACACUGUCCAAAGA (SEQ ID NO: 376); or UAUGGACACUGUCCAAAGAG (SEQ ID NO: 377). In some embodiments of the methods, the subject sustains at least a 24 hour fasting period after treatment.

Another aspect provides a vector comprising any of the polynucleotides as provided and described herein. In some embodiments, the vector is a viral vector. In some embodiments, the viral vector is a retroviral vector, adenoviral vector, lentiviral vector, herpesvirus vector, or adeno-associated viral vector (AAV).

Another aspect provides a composition including any of the fusion proteins or complexes as provided herein, any of the base editor systems as provided herein, any of the polynucleotides as provided herein, any of the cells as provided herein, or any of the vectors as provided herein. In some embodiments, the composition further includes a

pharmaceutically acceptable excipient, carrier, or vehicle. In some embodiments, the one or more guide polynucleotides and the fusion protein are formulated together or separately. In some embodiments, the composition further includes a ribonucleoparticle suitable for expression in a mammalian cell. In some embodiments, the composition further includes a lipid. In some embodiments, the composition comprises a lipid nanoparticle (LNP).

Yet another aspect provides a kit including any of the fusion proteins or complexes as provided herein, any of the base editor systems as provided herein, any of the polynucleotides as provided herein, any of the cells as provided herein, any of the vectors as provided herein, or any of the compositions as provided herein. In some embodiments, the kit further includes written instructions for the use of the kit in the treatment of Glycogen Storage Disease Type 1a (GSD1a).

Another aspect provides any of the fusion proteins or complexes as provided and described herein, any of the base editor systems as provided and described herein, any of the polynucleotides as provided and described herein, any of the cells as provided and described herein, any of the vectors as provided and described herein, or any of the compositions as provided and described herein, wherein the base editor comprises an mRNA sequence as set forth in SEQ ID NO: 396.

Another aspect provides any of the fusion proteins or complexes as provided and described herein, any of the base editor systems as provided and described herein, any of the polynucleotides as provided and described herein, any of the cells as provided and described herein, any of the vectors as provided and described herein, or any of the compositions as provided and described herein, wherein the base editor comprises a DNA sequence as set forth in SEQ ID NO: 397.

Another aspect provides any of the fusion proteins or complexes as provided and described herein, any of the base editor systems as provided and described herein, any of the polynucleotides as provided and described herein, any of the cells as provided and described herein, any of the vectors as provided and described herein, or any of the compositions as provided and described herein, wherein the base editor comprises an amino acid sequence as set forth in SEQ ID NO: 398.

In another aspect, a modified guide RNA (gRNA) comprising modified nucleotides is provided, wherein the gRNA comprises from 5' to 3' a polynucleotide sequence selected from the group consisting of:

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACU AAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 404);

UUUCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCU ACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 405);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAG GCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 406);

CCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUAC UAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 407);

ACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 408);

CACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCU ACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 409);

CCACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUC UACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 410);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAGAAAUACAGAAUCUACUAAAA CAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 411);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGCGGAAACGCAGAAUCUACUAAAA CAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 412);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCGAAAGAAUCUACUAAAACAAGGCAA AAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 413);

CAGUAUGGACACUGUCCAAAGUUUUAGUACCCGAAAGCAUCUACUAAAACAAGGCAA AAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 414); and

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACU AAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 415).

In some embodiments, the guide comprises at least about 50%-75% modified nucleotides. In some embodiments, the guide comprises at least about 85% or more modified nucleotides. In some embodiments, at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified. In some embodiments, at least about 3-5 contiguous nucleotides at each of the 5' and 3' termini of the gRNA are modified. In some embodiments, at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 50% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some

embodiments, at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 100 of the nucleotides present in a direct repeat or anti-direct repeat are modified. In some embodiments, at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified. In some embodiments, at least about 50% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified. In some embodiments, the guide comprises a variable length protospacer. In some embodiments, the guide comprises a 20-40 nucleotide protospacer. In some embodiments, the guide comprises a protospacer comprising at least about 20-25 nucleotides or at least about 30-35 nucleotides. In some embodiments, the protospacer comprises modified nucleotides. In some embodiments, the guide comprises two or more of the following:

at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified;

at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified;

at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified;

at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified;

a variable length protospacer; and

a protospacer comprising modified nucleotides.

In some embodiments of the modified gRNA, the gRNA comprises one or more modifications selected from the group consisting of 2′-*O*-methyl (2′-OMe), phosphorothioate (PS), 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), and constrained ethyl (S-cEt). In embodiments, the gRNA comprises 2′-*O*-methyl or phosphorothioate modifications. In an embodiment, the gRNA comprises 2′-*O*-methyl and phosphorothioate modifications. In an embodiment, the modifications increase base editing by at least about 2 fold.

In another aspect, a modified guide RNA (gRNA) is provided, wherein the gRNA comprises a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUmAmGmUACUCm UGmUmAmAmUGmAAAmAmUmUmACmAGAAUCUACmUmAAAACAAGGCAAmA

AUGmCCmGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGm GmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404);

mUsmUsmUsCAGmUAUmGmGmACAmCUGUCCAAAmGUUUUmAmGmUAC UCmUGmUmAmAmUGmAAAmAmUmUmACmAGAAUCUACmUmAAAACAAGGCA AmAAUGmCCmGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmU mGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 405);

mCsmAsGsmUAUmGmGmAmCAmCmUGUCCAAmAmGUmUUmUmAmGmUA CUmCmUmGmUmAmAmUGmAmAmAmAmUmUmACmAmGAAmUCUACmUmAmA AACAAGmGCAAmAAUGmCmCmGUGmUmUmUmAmUmCmUmCmGmUmCmAmAm CmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsGsmUmAUmGmGmAmCAmCUGUCCAAmAmGUUUmUAmGmUACU mCmUmGmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAAmUCUACUmAmA AACAAmGmGmCmAmAmAAUmGmCmCGUGmUmUmUmAmUmCmUmCmGmUmC mAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404); or

mCsmAsGsmUmAUmGmGmAmCAmCmUGUCmCmAAmAmGmUmUmUmUmA mGmUAmCmUmCmUmGmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAmAm UCUACmUmAmAmAAmCAmAmGmGmCmAmAmAAUmGmCmCmGmUGmUmUmU mAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsm UsmUsmU (SEQ ID NO: 404).

In another aspect, a modified guide RNA (gRNA) is provided, which comprises a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGmUmAm AmUmGmAmAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGC CGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGmUmAm AmUmGmAmAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGC CGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGm AmGmAmUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUAGUACmUmCmU mGmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGC AAAAUGCCmGUGUmUmUAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmG mGmCmGmAmGmAUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGGmAmA mAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCGUGUmUmUmAmUmCmU mCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 406); or

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUAGUACmUmCmU mGmGmAmAmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCmGUGUmUm UAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAUsmUs mUsmU (SEQ ID NO: 406).

In another aspect, a modified guide RNA (gRNA) is provided, which comprises a nucleic acid sequence, from 5' to 3', selected from

mCsmCsmAsGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAA AUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUU GUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 407);

mAsmCsmCsAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAA AAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACU UGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 408);

mCsmAsmCsCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGA AAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAAC UUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 409); or

mCsmCsmAsCCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUG AAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAA CUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 410).

In another aspect, a modified guide RNA (gRNA) is provided, which comprises a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAGAAAUAC AGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGG CGAGAUsmUsmUsmU (SEQ ID NO: 411);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGCG GAAA CGCAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGU UGGCGAGAUsmUsmUsmU (SEQ ID NO: 412);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGGAAACAGAA UCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAG AUsmUsmUsmU (SEQ ID NO: 406);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCGAAAGAAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmU smUsmU (SEQ ID NO: 413); or

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACCCGAAAGCAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmU smUsmU (SEQ ID NO: 414).

In another aspect, a modified guide RNA (gRNA) is provided, wherein the gRNA comprises a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAA UUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUG UUGGCGAGAUsmUsmUsmU (SEQ ID NO: 415); or

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAA UUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUG UUGGCGAGAmUsmUsmUsU (SEQ ID NO: 415).

In the above-delineated modified gRNAs, the “m” denotes a 2'-O-methyl and “s” denotes a phosphorothioate.

In another aspect, a formulation is provided which comprises a lipid nanoparticle comprising an mRNA expressing a base editor and a gRNA, wherein the base editor comprises a Cas9 domain and at least one adenosine deaminase variant comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N with reference to SEQ ID NO: 1:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase; and the gRNA comprises CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370).

In another aspect, a formulation is provided which comprises a lipid nanoparticle comprising an mRNA expressing a base editor, wherein the base editor comprises a Cas9 domain and at least one adenosine deaminase variant comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N with reference to SEQ ID NO: 1:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP

GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase; and a gRNA comprising CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370).

In an embodiment of the above-delineated formulations, the adenosine deaminase variant domain comprises the following combination of alterations I76Y + V82G + Y147D + F149Y + Q154S + D167N of SEQ ID NO: 1, or corresponding alterations in another adenosine deaminase. In some embodiments, the gRNA comprises 2′-*O*-methyl and/or phosphorothioate modifications. In some embodiments, the gRNA comprises 2′-*O*-methyl and phosphorothioate modifications. In some embodiments, the mRNA comprises one or more pseudouridines. In some embodiments, the mRNA comprises an N1-methylpseudouridine (m1Ψ).

Other aspects as described herein provide modified guide RNA sequences (gRNAs), (e.g., heavily modified gRNA sequences or "heavy mods"), such as described in Example 5 and as set forth in SEQ ID NOS: 404-415.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the pertinent art. The following references provide one of skill with a general definition of many of the terms used in this disclosure and the embodiments described herein: Singleton *et al.*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991).

In this application, the use of the singular includes the plural unless specifically stated otherwise. It must be noted that, as used in the specification, the singular forms "a," "an," and "the" include plural references unless the context clearly dictates otherwise. In this application, the use of "or" means "and/or," unless stated otherwise, and is understood to be inclusive. Furthermore, use of the term "including" as well as other forms, such as "include," "includes," and "included," is not limiting.

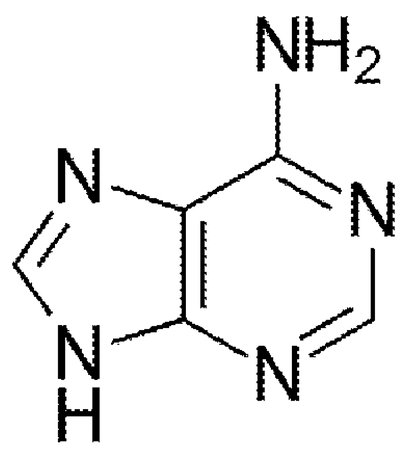
As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include")

or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. Any embodiments specified as "comprising" a particular component(s) or element(s) are also contemplated as "consisting of" or "consisting essentially of" the particular component(s) or element(s) in some embodiments. It is contemplated that any embodiment discussed in this specification can be implemented with respect to any method or composition of the present disclosure, and vice versa. Furthermore, compositions of the present disclosure can be used to achieve methods of the present disclosure.

The term "about" or "approximately" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art. Such acceptable range will depend in part on how the value is measured or determined, *i.e.*, the limitations of the measurement system.

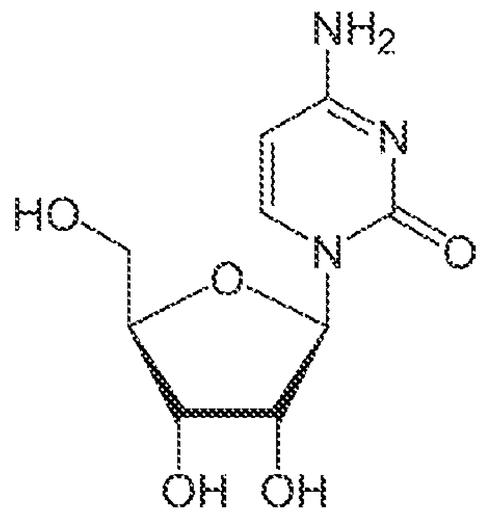
Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

Reference in the specification to "some embodiments," "an embodiment," "one embodiment" or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the present disclosures.

By "adenine" or " 9*H*-Purin-6-amine" is meant a purine nucleobase with the molecular formula $C_5H_5N_5$, having the structure 

, and corresponding to CAS No. 73-24-5.

By "adenosine" or " 4-Amino-1-[(2*R*,3*R*,4*S*,5*R*)-3,4-dihydroxy-5-(hydroxymethyl)oxolan-2-yl]pyrimidin-2(1*H*)-one" is meant an adenine molecule attached to

a ribose sugar via a glycosidic bond, having the structure , and corresponding to CAS No. 65-46-3. Its molecular formula is $C_{10}H_{13}N_5O_4$.

By "adenosine deaminase" or "adenine deaminase" is meant a polypeptide or fragment thereof capable of catalyzing the hydrolytic deamination of adenine or adenosine. In some embodiments, the deaminase or deaminase domain is an adenosine deaminase catalyzing the hydrolytic deamination of adenosine to inosine or deoxy adenosine to deoxyinosine. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenine or adenosine in deoxyribonucleic acid (DNA). The adenosine deaminases (*e.g.* engineered adenosine deaminases, evolved adenosine deaminases) provided herein may be from any organism (e.g., eukaryotic, prokaryotic), including but not limited to algae, bacteria, fungi, plants, invertebrates (e.g., insects), and vertebrates (e.g., amphibians, mammals). In some embodiments, the adenosine deaminase is an adenosine deaminase variant with one or more alterations and is capable of deaminating both adenine and cytosine in a target polynucleotide (*e.g.*, DNA, RNA). In some embodiments, the target polynucleotide is single or double stranded. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in DNA. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in single-stranded DNA. In some embodiments, the adenosine deaminase variant is capable of deaminating both adenine and cytosine in RNA.

By "adenosine deaminase activity" is meant catalyzing the deamination of adenine or adenosine to guanine in a polynucleotide. In some embodiments, an adenosine deaminase variant as provided herein maintains adenosine deaminase activity (e.g., at least about 30%, 40%, 50%, 60%, 70%, 80%, 90% or more of the activity of a reference adenosine deaminase (e.g., TadA*8.20 or TadA*8.19)).

By "Adenosine Base Editor (ABE)" is meant a base editor comprising an adenosine deaminase.

By "Adenosine Base Editor 8 (ABE8) polypeptide" or "ABE8" is meant a base editor as defined herein comprising an adenosine deaminase variant comprising an alteration at

amino acid position 82 and/or 166 of the following reference sequence: MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1). In some embodiments, ABE8 comprises further alterations, as described herein, relative to the reference sequence.

By "Adenosine Base Editor 8 (ABE8) polynucleotide" is meant a polynucleotide encoding an ABE8 polypeptide.

By "Adenosine Deaminase polynucleotide" is meant a polynucleotide encoding an adenosine deaminase polypeptide. In particular embodiments, the adenosine deaminase polynucleotide encodes an adenosine deaminase variant comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N. In some embodiments, the adenosine deaminase polynucleotide encodes an adenosine deaminase variant comprising one of the following combinations of alterations: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; or L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N.

In some embodiments, the deaminase or deaminase domain is a variant of a naturally occurring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase or deaminase domain does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, or at least 99.9% identical to a naturally occurring deaminase. In some embodiments, the adenosine deaminase is from a bacterium, such as, *E. coli*, *S. aureus*, *B. subtilis*, *S. typhi*, *S. putrefaciens*, *H. influenzae*, *C. crescentus*, or *G. sulfurreducens*.

In some embodiments, the adenosine deaminase is a TadA deaminase. In some embodiments, the TadA deaminase is an *E. coli* TadA (ecTadA) deaminase or a fragment thereof. In some embodiments, the ecTadA deaminase is truncated ecTadA. For example,

the truncated ecTadA may be missing one or more N-terminal amino acids relative to a full-length ecTadA. In some embodiments, the truncated ecTadA may be missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length ecTadA. In some embodiments, the truncated ecTadA may be missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length ecTadA. In some embodiments, the ecTadA deaminase does not comprise an N-terminal methionine. In some embodiments, the TadA deaminase is an N-terminal truncated TadA. In particular embodiments, the TadA is any one of the TadAs described in PCT/US2017/045381, which is incorporated herein by reference in its entirety.

In some embodiments, the TadA deaminase is TadA variant. In some embodiments, the TadA variant is TadA*7.10 comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N. In some embodiments, the TadA variant is TadA*7.10 comprising a combination of alterations selected from among the following: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; or L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N. In some embodiments, the TadA variant is MSP605, MSP680, MSP823, MSP824, MSP825, MSP827, MSP828, or MSP829.

"Administering" is referred to herein as providing one or more compositions described herein to a patient or a subject.

By "agent" is meant any small molecule chemical compound, antibody, nucleic acid molecule, or polypeptide, or fragments thereof.

By "alteration" is meant a change (*e.g.* increase or decrease) in the structure, expression levels or activity of a gene or polypeptide as detected by standard art known methods such as those described herein. As used herein, an alteration includes a change in a polynucleotide or polypeptide sequence or a change in expression levels, such as a 10% change, a 25% change, a 40% change, a 50% change, or greater.

By "ameliorate" is meant decrease, suppress, attenuate, diminish, arrest, or stabilize the development or progression of a disease.

By "analog" is meant a molecule that is not identical, but has analogous functional or structural features. For example, a polynucleotide or polypeptide analog retains the biological activity of a corresponding naturally-occurring polynucleotide or polypeptide,

while having certain modifications that enhance the analog's function relative to a naturally occurring polynucleotide or polypeptide. Such modifications could increase the analog's affinity for DNA, efficiency, specificity, protease or nuclease resistance, membrane permeability, and/or half-life, without altering, for example, ligand binding. An analog may include an unnatural nucleotide or amino acid.

By "base editor (BE)," or "nucleobase editor polypeptide (NBE)" is meant an agent that binds a polynucleotide and has nucleobase modifying activity. In various embodiments, the base editor comprises a nucleobase modifying polypeptide (*e.g.*, a deaminase) and a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9 or Cpf1) in conjunction with a guide polynucleotide (*e.g.*, guide RNA (gRNA)). Representative nucleic acid and protein sequences of base editors are provided in the Sequence Listing as SEQ ID NOs: 2-11 and 378.

By "base editing activity" is meant acting to chemically alter a base within a polynucleotide. In one embodiment, a first base is converted to a second base. In one embodiment, the base editing activity is adenosine or adenine deaminase activity, *e.g.*, converting A•T to G•C.

In some embodiments, base editing activity is assessed by efficiency of editing. Base editing efficiency may be measured by any suitable means, for example, by sanger sequencing or next generation sequencing. In some embodiments, base editing efficiency is measured by percentage of total sequencing reads with nucleobase conversion effected by the base editor, for example, percentage of total sequencing reads with target A•T base pair converted to a G•C base pair or target C•G base pair to a T•A base pair. In some embodiments, base editing efficiency is measured by percentage of total cells with nucleobase conversion effected by the base editor, when base editing is performed in a population of cells.

The term "base editor system" refers to an intermolecular complex for editing a nucleobase of a target nucleotide sequence. In various embodiments, the base editor (BE) system comprises (1) a polynucleotide programmable nucleotide binding domain, a deaminase domain (*e.g.*, cytidine deaminase or adenosine deaminase) for deaminating nucleobases in the target nucleotide sequence; and (2) one or more guide polynucleotides (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In various embodiments, the base editor (BE) system comprises a nucleobase editor domain selected from an adenosine deaminase or a cytidine deaminase, and a domain having

nucleic acid sequence specific binding activity. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable DNA binding domain and a deaminase domain for deaminating one or more nucleobases in a target nucleotide sequence; and (2) one or more guide RNAs in conjunction with the polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the base editor is a cytidine base editor (CBE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE). In some embodiments, the base editor is an adenine or adenosine base editor (ABE) or a cytidine or cytosine base editor (CBE).

The term "Cas9" or "Cas9 domain" refers to an RNA guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease.

The term "Cas9" or "Cas9 domain" refers to an RNA guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a Casnl nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat) associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3′-5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA," or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.*, Jinek M., *et al. Science* 337:816-821(2012), the entire contents of which is hereby

incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of Streptococcus pyogenes." Ferretti *et al.*, Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., *et al.*, Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., *et al.*, *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) RNA Biology 10:5, 726-737; the entire contents of which are incorporated herein by reference.

A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9) or catalytically inactive Cas9. Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek *et al., Science.* 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell.* 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al., Science.* 337:816-821(2012); Qi *et al.*, *Cell.* 28;152(5):1173-83 (2013)). In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity. In some embodiments, a dCas9 domain comprises D10A and an H840A mutation or corresponding mutations in another Cas9. In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an "nCas9" protein (for "nickase" Cas9). It should be appreciated that

additional Cas9 proteins (*e.g.*, a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9), including variants and homologs thereof, are within the scope of this disclosure. Exemplary Cas9 proteins include, without limitation, those provided herein. In some embodiments, the Cas9 protein is a nuclease dead Cas9 (dCas9). In some embodiments, the Cas9 protein is a Cas9 nickase (nCas9). In some embodiments, the Cas9 protein is a nuclease active Cas9.

In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to wild-type Cas9. In some embodiments, the Cas9 variant may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 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 or more amino acid changes compared to wild-type Cas9. In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild-type Cas9. In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild-type Cas9.

In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length.

In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs:

NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1), *Campylobacter jejuni* (NCBI Ref: YP_002344900.1) or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any other organism.

In some embodiments, the Cas9 is from *Neisseria meningitidis* (Nme). In some embodiments, the Cas9 is Nme1, Nme2 or Nme3. In some embodiments, the PAM-interacting domains for Nme1, Nme2 or Nme3 are N_4GAT, N_4CC, and N_4CAAA, respectively (*see e.g.*, Edraki, A., *et al.*, A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for *In Vivo* Genome Editing, Molecular Cell (2018)).

In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas9 protein, *e.g.*, one of the Cas9 sequences provided herein. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only one or more fragments thereof. For example, in some embodiments, a Cas9 fusion protein provided herein comprises a Cas9 fragment, wherein the fragment binds crRNA and tracrRNA or sgRNA, but does not comprise a functional nuclease domain, *e.g.*, in that it comprises only a truncated version of a nuclease domain or no nuclease domain at all. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of skill in the art.

In some embodiments, Cas9 refers to a Cas9 from archaea (*e.g.* nanoarchaea), which constitute a domain and kingdom of single-celled prokaryotic microbes. In some embodiments, Cas9 refers to CasX or CasY, which have been described in, for example, Burstein *et al.*, "New CRISPR-Cas systems from uncultivated microbes." Cell Res. 2017 Feb 21. doi: 10.1038/cr.2017.21, the entire contents of which is hereby incorporated by reference. Using genome-resolved metagenomics, a number of CRISPR-Cas systems were identified, including the first reported Cas9 in the archaeal domain of life. This divergent Cas9 protein was found in little- studied nanoarchaea as part of an active CRISPR-Cas system. In bacteria, two previously unknown systems were discovered, CRISPR-CasX and CRISPR-CasY, which are among the most compact systems yet discovered. In some embodiments, Cas9 refers to CasX, or a variant of CasX. In some embodiments, Cas9 refers to a CasY, or a variant of

CasY. It should be appreciated that other RNA-guided DNA binding proteins may be used as a nucleic acid programmable DNA binding protein (napDNAbp), and are within the scope of this disclosure.

In particular embodiments, napDNAbps useful in the methods described herein include circular permutants, which are known in the art and described, for example, by Oakes *et al.*, Cell 176, 254–267, 2019.

"Co-administration" or "co-administered" refers to administering two or more therapeutic agents or pharmaceutical compositions during a course of treatment. Such co-administration can be simultaneous administration or sequential administration. Sequential administration of a later-administered therapeutic agent or pharmaceutical composition can occur at any time during the course of treatment after administration of the first pharmaceutical composition or therapeutic agent.

The term "conservative amino acid substitution" or "conservative mutation" refers to the replacement of one amino acid by another amino acid with a common property. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms (Schulz, G. E. and Schirmer, R. H., Principles of Protein Structure, Springer-Verlag, New York (1979)). According to such analyses, groups of amino acids can be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz, G. E. and Schirmer, R. H., supra). Non-limiting examples of conservative mutations include amino acid substitutions of amino acids, for example, lysine for arginine and vice versa such that a positive charge can be maintained; glutamic acid for aspartic acid and vice versa such that a negative charge can be maintained; serine for threonine such that a free –OH can be maintained; and glutamine for asparagine such that a free $–NH_2$ can be maintained.

The term "coding sequence" or "protein coding sequence" as used interchangeably herein refers to a segment of a polynucleotide that codes for a protein. Coding sequences can also be referred to as open reading frames. The region or sequence is bounded nearer the 5' end by a start codon and nearer the 3' end with a stop codon. Stop codons useful with the base editors described herein include the following:

Glutamine	CAG → TAG Stop codon
	CAA → TAA
Arginine	CGA → TGA

Tryptophan TGG → TGA

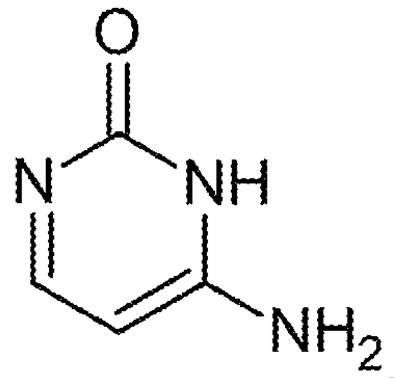
TGG → TAG

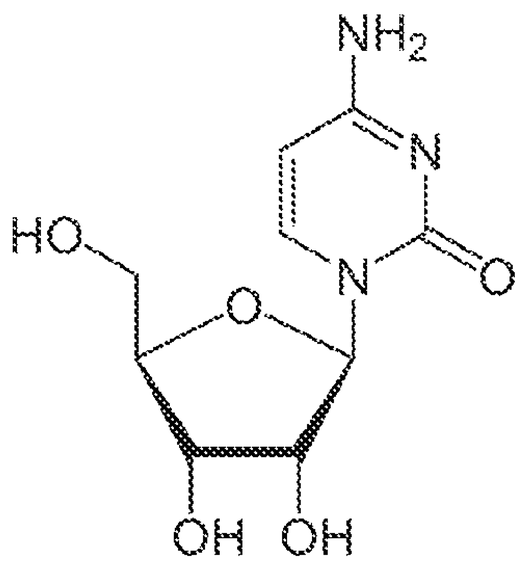
TGG → TAA

By "codon optimization" is meant a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (*e.g.* about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ (visited Jul. 9, 2002), and these tables can be adapted in a number of ways. See, Nakamura, Y., *et al.* "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In some embodiments, one or more codons (*e.g.* 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding an engineered nuclease correspond to the most frequently used codon for a particular amino acid.

By "complex" is meant a combination of two or more molecules whose interaction relies on inter-molecular forces. Non-limiting examples of inter-molecular forces include covalent and non-covalent interactions. Non-limiting examples of non-covalent interactions include hydrogen bonding, ionic bonding, halogen bonding, hydrophobic bonding, van der Waals interactions (e.g., dipole-dipole interactions, dipole-induced dipole interactions, and London dispersion forces), and π-effects. In an embodiment, a complex comprises polypeptides, polynucleotides, or a combination of one or more polypeptides and one or more polynucleotides. In one embodiment, a complex comprises one or more polypeptides that associate to form a base editor (e.g., base editor comprising a nucleic acid programmable

DNA binding protein, such as Cas9, and a deaminase) and a polynucleotide (e.g., a guide RNA). In an embodiment, the complex is held together by hydrogen bonds. It should be appreciated that one or more components of a base editor (e.g., a deaminase, or a nucleic acid programmable DNA binding protein) may associate covalently or non covalently. As one example, a base editor may include a deaminase covalently linked to a nucleic acid programmable DNA binding protein (e.g., by a peptide bond). Alternatively, a base editor may include a deaminase and a nucleic acid programmable DNA binding protein that associate noncovalently (e.g., where one or more components of the base editor are supplied in trans and associate directly or via another molecule such as a protein or nucleic acid). In an embodiment, one or more components of the complex are held together by hydrogen bonds.

By "cytosine" or " 4-Aminopyrimidin-2(1*H*)-one" is meant a purine nucleobase with the molecular formula $C_4H_5N_3O$, having the structure , and corresponding to CAS No. 71-30-7.

By "cytidine" is meant a cytosine molecule attached to a ribose sugar via a glycosidic bond, having the structure , and corresponding to CAS No. 65-46-3. Its molecular formula is $C_9H_{13}N_3O_5$.

By "Cytidine Base Editor (CBE)" is meant a base editor comprising a cytidine deaminase.

By "Cytidine Base Editor (CBE) polynucleotide" is meant a polynucleotide comprising a CBE.

By "cytidine deaminase" or "cytosine deaminase" is meant a polypeptide or fragment thereof capable of deaminating cytidine or cytosine. In one embodiment, the cytidine deaminase converts cytosine to uracil or 5-methylcytosine to thymine. The terms "cytidine

deaminase" and "cytosine deaminase" are used interchangeably throughout the application. Petromyzon marinus cytosine deaminase 1 (PmCDA1) (SEQ ID NO: 13-14), Activation-induced cytidine deaminase (AICDA) (SEQ ID NOs: 15-21), and APOBEC (SEQ ID NOs: 12-61) are exemplary cytidine deaminases. Further exemplary cytidine deaminase (CDA) sequences are provided in the Sequence Listing as SEQ ID NOs: 62-66 and SEQ ID NOs: 67-189.

By "cytosine" is meant a pyrimidine nucleobase with the molecular formula C4H5N3O.

By "cytosine deaminase activity" is meant catalyzing the deamination of cytosine or cytidine. In one embodiment, a polypeptide having cytosine deaminase activity converts an amino group to a carbonyl group. In an embodiment, a cytosine deaminase converts cytosine to uracil (i.e., C to U) or 5-methylcytosine to thymine (i.e., 5mC to T). In some embodiments, a cytosine deaminase as provided herein has increased cytosine deaminase activity (e.g., at least 10-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold or more) relative to a reference cytosine deaminase.

The term "deaminase" or "deaminase domain," as used herein, refers to a protein or fragment thereof that catalyzes a deamination reaction. Exemplary deaminases include cytidine and adenosine deaminases.

"Detect" refers to identifying the presence, absence or amount of the analyte to be detected.

By "detectable label" is meant a composition that when linked to a molecule of interest renders the latter detectable, via spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes, magnetic beads, metallic beads, colloidal particles, fluorescent dyes, electron-dense reagents, enzymes (for example, as commonly used in an enzyme-linked immunosorbent assay (ELISA)), biotin, digoxigenin, or haptens.

By "disease" is meant any condition or disorder that damages or interferes with the normal function of a cell, tissue, or organ. An example of a disease includes Glycogen Storage Disease Type 1 (also known as GSD1 or Von Gierke Disease). In some embodiments, the GSD1 is Type 1a (GSD1a).

The term "effective amount," as used herein, refers to an amount of a biologically active agent that is sufficient to elicit a desired biological response. In some embodiments, an effect amount is an amount required to ameliorate the symptoms of a disease relative to an

untreated patient. The effective amount of an active agent(s) used to practice therapeutic methods and treatment of a disease varies depending upon the manner of administration, the age, body weight, and general health of the subject. Ultimately, the attending physician or veterinarian will decide the appropriate amount and dosage regimen. Such amount is referred to as an "effective" amount. In one embodiment, an effective amount is the amount of a base editor described herein (*e.g.*, a fusion protein comprising a programable DNA binding protein, a nucleobase editor and gRNA) sufficient to introduce an alteration in a gene of interest (*e.g.*, G6PC) in a cell (*e.g.*, a cell *in vitro*, *in vivo*, or *ex vivo*). In one embodiment, an effective amount is the amount of a base editor required to achieve a therapeutic effect (*e.g.*, to reduce or control GSD1a or a symptom or condition thereof). Such therapeutic effect need not be sufficient to alter G6PC in all cells of a subject, tissue or organ, but only to alter G6PC in about 1%, 5%, 10%, 25%, 50%, 75% or more of the cells present in a subject, tissue or organ. In one embodiment, an effective amount is sufficient to ameliorate one or more symptoms of GSD1a.

In some embodiments, an effective amount of a fusion protein provided herein, *e.g.*, of a nucleobase editor comprising a nCas9 domain and a deaminase domain (*e.g.*, adenosine deaminase) refers to the amount of the fusion protein that is sufficient to induce editing of a target site specifically bound and edited by the nucleobase editors described herein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, a nuclease, a hybrid protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, *e.g.*, on the specific allele, genome, or target site to be edited, on the cell or tissue being targeted, and/or on the agent being used.

By "fragment" is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

By "glucose-6-phosphatase (G6PC) polypeptide" is meant a polypeptide or fragment thereof having at least about 95% amino acid sequence identity to NCBI Accession No. AAA16222.1. A wild-type G6PC polypeptide is capable of catalyzing the hydrolysis of D-glucose 6-phosphate to D-glucose and orthophosphate, while a G6PC polypeptide comprising a deleterious mutation lacks or has reduced catalytic activity. In particular embodiments, a

method of editing a G6PC polynucleotide is provided, in which the method comprises a single nucleotide polymorphism (SNP) associated with Glycogen Storage Disease Type 1a (GSD1a). In one embodiment, the A•T to G•C alteration at the SNP associated with GSD1a changes a glutamine (Q) to a non-glutamine (X) amino acid in the G6PC polypeptide. In another embodiment, the A•T to G•C alteration at the SNP associated with GSD1a changes an arginine (R) to a non-arginine (X) in the G6PC polypeptide. In one embodiment, the SNP associated with GSD1a results in expression of an G6PC polypeptide having a non-glutamine (X) amino acid at position 347 or a non-arginine (X) amino acid at position 83. In one embodiment, the base editor correction replaces the glutamine at position 347 with a non-glutamine amino acid (X). In another embodiment, the base editor correction replaces the arginine at position 83 with a non-arginine amino acid (X). Mutations associated with GSD1a are known in the art and described, for example, in Chou *et al*., Hum. Mutat. 29:921-930, 2008, which is incorporated herein by reference. Methods for detecting glucose-6-phosphatase activity are known in the art and described, for example by Varga *et al*., Int. J. Mol. Sci. 2019, 20, 5039, which is incorporated herein by reference.

In particular embodiments, G6PC comprises one or more alterations relative to the following reference sequence. In particular embodiments, G6PC associated with GSD1a comprises one or more mutations selected from Q347X and R83C. An exemplary G6PC amino acid sequence from *Homo Sapiens* is provided below:

```
1    MEEGMNVLHD FGIQSTHYLQ VNYQDSQDWF ILVSVIADLR NAFYVLFPIW FHLQEAVGIK
61   LLWVAVIGDW LNLVFKWILF GQRPYWWVLD TDYYSNTSVP LIKQFPVTCE TGPGSPSGHA
121  MGTAGVYYVM VTSTLSIFQG KIKPTYRFRC LNVILWLGFW AVQLNVCLSR IYLAAHFPHQ
181  VVAGVLSGIA VAETFSHIHS IYNASLKKYF LITFFLFSFA IGFYLLLKGL GVDLLWTLEK
241  AQRWCEQPEW VHIDTTPFAS LLKNLGTLFG LGLALNSSMY RESCKGKLSK WLPFRLSSIV
301  ASLVLLHVFD SLKPPSQVEL VFYVLSFCKS AVVPLASVSV IPYCLAQVLG QPHKKSL
```

(SEQ ID NO: 379)

By "glucose-6-phosphatase (G6PC) polynucleotide" is meant a nucleic acid molecule encoding an G6PC polypeptide, as well as the introns, exons, and regulatory sequences associated with its expression, or fragments thereof. In embodiments, an G6PC polynucleotide is the genomic sequence, mRNA, or gene associated with and/or required for G6PC expression. An exemplary G6PC nucleotide sequence from *Homo Sapiens* is provided below (GenBank: U01120.1):

```
  1 ATAGCAGAGC AATCACCACC AAGCCTGGAA TAACTGCAAG GGCTCTGCTG ACATCTTCCT
 61 GAGGTGCCAA GGAAATGAGG ATGGAGGAAG GAATGAATGT TCTCCATGAC TTTGGGATCC
```

```
121 AGTCAACACA TTACCTCCAG GTGAATTACC AAGACTCCCA GGACTGGTTC ATCTTGGTGT
181 CCGTGATCGC AGACCTCAGG AATGCCTTCT ACGTCCTCTT CCCCATCTGG TTCCATCTTC
241 AGGAAGCTGT GGGCATTAAA CTCCTTTGGG TAGCTGTGAT TGGAGACTGG CTCAACCTCG
301 TCTTTAAGTG GTAAGAACCA TATAGAGAGG AGATCAGCAA GAAAAGAGGC TGGCATTCGC
361 TCTCGCAATG TCTGTCCATC AGAAGTTGCT TTCCCCAGGC TATTCAGGAA GCCACGGGCT
421 ACTCATGCTT CCAACCCCTC TCTCTGACTT TGGATCATCT ACATAAAGGG GGAAGACAGA
481 AAAAATCCTA CCAGTGAGTT GAAAATACAG GAAAGCCTAT TTCATATGGG TTAAAGGGTA
541 GGACAGTTGA ATTTCGTGAA AAGTCTGAGT TATATAGGCT TTGAGCAAAG AGTTTTATTA
601 GTATGAAGCA GAAGAGGTAA CATAAAGAAA GATGTATGGG GCCAGGCATG GTGGCTCACA
661 CCTGTAATCC CAGCACTTTG GGAGGCCGAG GTGGGCGAAT CACTCCTGGG TGAACTCAGG
721 AGTTCAAGAC CAGCCTGGGC AACATGGCGA AACTCCATCT CTACAAAAAC ATTACGAAAA
781 TTAGCTGGGC GTGTTGGTGC TGTAGTCCCA GCTACTCAGG AGGCTGAGGT GAGAGGCGGA
841 GGAGGTTGCA GTGAGTCAAG ATCATGCCAC TGCACTCCAG CCTGGGCAAC AGAGTAAGAC
901 CCTGTCTCAA AAAAAAAAAA AAGATAGATG ATGTATGCTG TATGAAAAAA GGAAACACAC
961 AGATGATTCA ACAGCCTGTT TTGTGGGGTA ATGAAAAGTC ACCCTGGGAA CTGGGCTCCA
1021 GCCCTCGTTC TGCCACCCAC CAACTACATG TCCTTGGCAA GTCATATCAA TTATCTGAGT
1081 TTCTGTTTTA TAATCTACAA ATAGGTTATC TCTGGCAGCT TAATAATAAT CAGGGTTAAC
1141 ATTTATTAAA CAGTGTGTGC CAGTCCATGT GCTATGTGCT TTTCTGTGAG GTAGTTACTG
1201 CTATTTACAG AAACAGTAGA TGCAGAGACC AAGGTGCTGA GTTAAATGAT TAGGCCAACA
1261 AGGTTAGTAC ATGCCGAGCC AGGATGGAAG CCCAGGTAGG CAGGCTGGCT TCCGCGGCAA
1321 TGCTCTTATG AACTATGTTA CGTCCAGTGC TGATAAACTG ACTCTCTGGG GAGCAGGGGA
1381 AAGCCCTGAG TTTAGCATTT GCCAATTTCT ATCACGTAAA CATTCCCATT CTGGCCACTT
1441 TCTTTCTTTC TTTCTTTTGT TTGTTTGTTT GAGATGGAGT CTCGCACTGT TGCCTGGCTG
1501 GAGTGCAATG GTGCAATCTC AGCTCACTGC AACCTCTGCC TCTCCGGTTC AAGTGATTCT
1561 CCTGCCTCAG CCTCCCAAGT AGCTGGGATT ACAGGTGCCC GCCACCATGC CCAGCTAATT
1621 TTTTTTGTAT TTTTAGTAGA GACATGGTTT CACTATGTTG ACTAGGCTGG TCTCGAACTC
1681 CTGACCTCAT GATCTGCCTG CCTTGGCCTC CCTAAGTGCT AGGATTACAG GCGTGAGCCA
1741 CTACACCCAG CCGCATGATT CTAAAAAATA AAAAGATGAA GTGTTATTCC AAACATCTGA
1801 TCTCCATTGA AGAACCATGC AATCTCTCTG GGTTGATAGA GGCCAGAGTT AGTGGCTCTC
1861 CCTGATTTCG GTGAGAAATC ACTATTCCAC CATCACGGGA TAAAAGGCAT CCTGACTGGC
1921 GGTTGACACC TATTTCCACA GTGAAAGATA TATCTAGTAC TTTTAAAGGG GAAGTGGTTT
1981 GTCTGAGATA CTCTGTTTCA AAGTAGAGAG GATACAGAAC AAGCATCTGA AGCTATATAC
2041 ATCCTTACAG AGAGCAATTC TGATGGAAAT GCAGGCCATG TTTCCCTGGG GGGGGCTCGT
2101 CCTAGGGGCT GGAGTGCATT CTCTGATGTC AGAGGAAATG CAAGATTCCC TGAGGCCTGA
2161 GGGAACCCAT GGTATATGCA AGTCCAAGTT TCAAACTGTA GTTCCATATG CATTCTTCCA
2221 GGACAAATAC TTCTTGAGGT TAAAAAAAAA AAGTCACATA GCTGCCATTT TATGGATTTC
2281 AGGATTTTTT TTTTTTTTTT TTTGAGATGG AGTCTTGCTC TGTCACCCAG CCTGTAGTGC
2341 AGTGGCATAA TCTCGGCTCA CGGCAACCTC CGCCTCCCAG GTTCAAGCGA TTCTCTTGCC
2401 TTAGCCTCCC GAGTAGCTGG GATTACAGTC ACGCACCACC ACATCTGGCT AATTCTTTAT
2461 ATTTTTTGGT AGAAACGGTG TTTCACCATG TTGGCCAGGC TGGTCTCAAA CTCCTGACCT
```

2521 CATGTGATCT GCCTGCCTTG GCCTCCCAAA GTGCTGAGAT TACAGGTGTG AGCCACCGCG
2581 CCTGCCTGGA GTTCAGAATC TTGGGCTTCA TTATTTGTGT TTAAATAGAT CATACAGTCA
2641 GGCACGGTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG GCTGAGGTGG GAGGATTGCC
2701 TGAGTTCAGG AGATGGAGAC CAGCCTGGGC AACATGGTGA AACCCCGTCT CTACTAAAAA
2761 TACAAAAACT AGCTGGATGT GGTGGCACAC ACCTGTAGTC CCAGCTATTC AGGAGGCTGA
2821 GGTGGGAGGA TCCCAGGAGG TAGAGGTCAC AATGAGCCGA GATTGCGCCA CTGCACTCCA
2881 GGCTGGGTTA CTGAGCCAGA TCCTGTCTCA AAAAAAAAAA AGATAATACA TTCAAACAGT
2941 TCAAAATGCA AAAGTTACAT ACATAAGGAA GTGTCATGAA ATATCTCCCT CTCACACTTC
3001 TCCCCAGCCA CCCAGTTCTC CCTTCTAGAG GCAACATGTG AAATCCTTCT CAGGCTACAC
3061 TCTTCTTGAA GGTGTAGGCT TTGGGCAAAA GCATTCATTC AGTAACCCCA GAAACTTGTT
3121 CTGTTTTTCC ATAGGATTCT CTTTGGACAG CGTCCATACT GGTGGGTTTT GGATACTGAC
3181 TACTACAGCA ACACTTCCGT GCCCCTGATA AAGCAGTTCC CTGTAACCTG TGAGACTGGA
3241 CCAGGTAAGC GTCCCAGCCC CTGCAGACAG AAGCTGAGTG GACCTCGTTT ACCTGTTATG
3301 GATGAAACTG ACCTTGAGGG GACATGAGGA GAGCCATTCC TTTGTACTTT TGTCATGCTC
3361 TTCAATTGGC ACAAATTAAT TCACTTCTGC AATACTTTCC TGAATAGCAC AGTAGTATTG
3421 GAAATCTGCC TATTACAGAA CCTGGATGGA GTCCAGAGAG GCACGGGCAT CCATGGGCAA
3481 AGGGCTCGTG AGAGTCACCG CCCTGCAGCG CTGTGTCCTG AGAAAGGAGG GGGCAGAAGC
3541 CTGAGCTTCT GGGGGTCCTT CCCAATGGCC TGGCCCACTG GATGTGCCCT CCTGAGCTGA
3601 CCGTCCAATC CCTTGCCCTC TCTGTGCCTA CGTTTTATTA GTTACAGCCA GATGGTTACT
3661 GTCAAATCAA ATGATAGATT TCATTTTCAG TATGTAATAG GAAGCCCCTC CCTCACCCTA
3721 AAGTCTCAGC TGCCCTCTAA GACTAGTACT CTCTAAGGTA CTAGTATCCC TTCCTCAGAG
3781 ACCCTTTCCC TGACCCCAAA ACTAGGGAAG GTCCCTTAGT TATTTGCTCT CACAGACCAC
3841 GCATTTACCT CAGAGCATAT TCACTCATTC AGCTGTTACT TACCAAGCAC CTACTGGGAG
3901 CTATACACTG TTCTATGTGC TAGGGATACC TCTGTCAGTG AACAACACAG ACACAAAGAT
3961 CCCTGCCCTT GTGGAGCTGA AATCTGAATA GAGGAGGTGA AATATACAAA AATTATAATA
4021 AATAAGTAAA CTAGGCCAGT TGTGGTTGCT CATGCCTGTA ATCCCAGCAC TTTGGGAAGC
4081 CAAGGTAGGT AGATCACCTG AGGTCAGGAG TTCAAAACCA GCCTGGCCAA CATTGCAAAA
4141 TCCTGTCTTT ACTAAAAATG GAAAAATTGG TCAGGCGTGA TGGCACACGC CTGTAGTCTC
4201 AGCTACCTGG GAGGCTGAGG CAGGAGAATC GCTTGAACCT GGGAGGCAGA GGTTGCAGTG
4261 AACCGAGATC GGACCACTGC ACTCCAGCCT GAATGACAGA ACGAGACTCT GTCTCAAAAA
4321 AAAAGTAAAC TATTAATATG TAGGATAGGC CAGGCACGGT GGCTCACCCT GTAATCCCAG
4381 CACTTTGGGA GGCTGAGGCG GGTGGATCAC CTGAGGTGAG GAGTTCAAGA CCAGCCTGGC
4441 CAACATGGCA AAACCCTGTC TCTACTAAAA ATACAAAAAT TAGCTGGGTG TCCTGGTGCA
4501 TGCCTGTAAT CTGAGCTACT CAGGAGGCTA AGGCAGGAGA ATCGCTTGAA CCTGGGAGGT
4561 GGTGAGCCAA GATTGCGCCA TTGCACTCCA GCCTGGGCGA CAAAATGAGA CACCATCTGA
4621 AAAAAAAAAA AAAATATATA TATATATACA CACACACACA CACACACACA CACACACACA
4681 TATAATACTA GAAAATGATT GTTTATAGGC AAAAAAAAAA AAAAAGAAGA AGAAGAAGAA
4741 AAGGAAAGGA GAAGGAAAGA AGGACCAAAC ATCTTTTGTA GAAATATGTT TGCTTTCATC
4801 ATAACAGCTT GTTATCAAGG ATGAATTTCT CCCTGAAATT AATGGAGGCA CAGACTGGAA
4861 AGTTTAAAGT GGCTTTAAGA GGTTATTTTA TTTAGTCCTC TGTCTTAATA GAAGCAAATT

4921 ATTATCTCTG CTCCTTAGGT AGAGTAGCTA AGGCTCAGAA AGTAGGCCGG GCGCGGTGGC
4981 TCACGCCTGT AATCCTAGCA CTTTGGGAGG CCAACGCAGG TGGATCACCT GAGGTCAGGA
5041 GTTTGAGACC AGCCTGGCCA ACATGGTGAA ACCTCGTCAC TAATAAAAAA ATACAAAAAC
5101 TTAGCCAGGC ATGGTGGCGG GCGCCTGTAA TCCCAGCTAC CCAGGAGGCT GCGGCAGGAG
5161 AATCACTTCA ACCCGGGAGG CAGAGGTTGC AGTGAGCTGA AATCACACCA CTGCACTCCA
5221 GCCTTGGTGA CAGAGAAAGA TTCTGTCAGG AAAAAAAAAA AAAAGTTTAA ATGAATTACC
5281 CAAGGTATAT AATTGTTAGT GTTAGAAGGA AGAAGAAGGG AGGGAGGAAG GAAGGGAGAA
5341 AGAAAGGGAA GGAGGAAGGG AGGGAGGGAA GAAAGCCTTT ATTTATCTAT GGGGTTCCCT
5401 GGAAAGCAGG CTGAAATGGA GATTCACGTG CAGGAGTTTA GATACTCTGG GGAACTATAC
5461 TTGTAGAAGG GAAGGAACAG GAACAGGGCA GAAGGAGAGG TCCGGTTGTG ATTCTGCCTC
5521 ATCCAACCCC ACAGCGAGCT CTGAAGCTGG GGATGGCTCC TCAGAGTTGG TCCAAGTTGG
5581 GACAAGGGAA TCAGACCCTG GGGAGAGCGT AACCTTGATC AAGGCGACTC TCTTTAGCCC
5641 AGGGCAATGC CAGGAGAAGG CTGAGAGCAG AAAGCCATCT ACCATCACAC TCTCAACAGC
5701 TACGAAATAA GTCCTGCAGT TCAGGAGGGA GGTCTGGGCG GCACATCTCA GGACCCTCTA
5761 TCTCTCAGGG TAGAGGAATT AAGAATGGGA TGGGAACCAG ACGGGCCATG GTGGCTCACA
5821 CCTATAATCC CAACACTTTG GGAGGCCAAG GGTAGGAGGA TTGCTTGAGC CCAAGAGTTC
5881 AAAACCAGCC TGGGCAAAAA CAATCAAACA AACAAACAAA ACACATTTAA AAAATTTGCT
5941 GTGTGTGGTG GTGTGCACCT GTGGTCCCAG CTACTCAGGG GGCTGAGGTG GGAGGATTGC
6001 TTGAGTCCAG GAGGTCGAGG CTGCAGTGAG CTATGATCAT GGCACTGCAT TGCAGCCTAG
6061 GAGACAAAGC AAGACACTGT CTCTAAAAAA ACAAAAAACA AACAAATAAA AAAACGGAAC
6121 CGGTTGCAAG CAGGGTTAAA TAGCGTGGTC AGAGTAGGAC TCACTGAGAA TATGAGATCT
6181 GAGTCAAGTC TTCAAGGATG TGAGGAAGTA AGTTTCTGGC AGAAGAGCTG TGAAGGGCTG
6241 TCTGGCCAGA GAAGATTGCA ATGCAAAAGC CCTGAGGTGG GAACGTGTTT GGTGTGTTTA
6301 AAGGAAAGCA ATGAGGCCAG TGTAGCCAGA ACAGAGTGTG CAAGGAGAGA AGGAACAGAA
6361 GATGTGGAGG GCAGATCAGT TTGTAATTGT ACGCCCAGTA TGCTGATTCT TTGTGTAATC
6421 TCCAGACTGT ATTAAACTGC AAGAGCAGGG CCCCTCTCTG GCTTTGCTCA TCATTGTATT
6481 CCCAGAGCCT TGCACAATGC TTGGTGCATA GGAGATGGAA ATTTGTTAAA TAAATGAATT
6541 ATGGATAACG AATGGATGGT AAGATGGGTG GATGGATGGG GGGTGAACGG ATGGATGGGG
6601 GGTGAATGGA TGGATGAATG GGTAGATGGG TGGATAGGGG GATGGCTGGG TGGCTGGGTA
6661 GATGATGCAC TGTCTCCCAG ATGAGGACCT TTTCACCTTT ACTCCATTCT CTTTCCTGCC
6721 CTTTAGGGAG CCCCTCTGGC CATGCCATGG GCACAGCAGG TGTATACTAC GTGATGGTCA
6781 CATCTACTCT TTCCATCTTT CAGGGAAAGA TAAAGCCGAC CTACAGATTT CGGTAAGAAC
6841 TCACCACTGG GGTGTAGGTG GTGGAGGGCA GGAGGCAGCT CTCTCTGTAG CTGACACACC
6901 ACGTATTCTT CCTCACATCC CCCTAGCCCG CTCCCACACC TGGGCAGCCG CTGATTAAGA
6961 GTTGTGGCAC TTTGGATAGG GATAAACCTC AGAGTCAGGG AATGTTTGGG CTGAAAGGGA
7021 TCCAGTAGTG CAATCCGTTG TTTTACAGAT AAGGAAACAA AGCCCAACAC CATGAAGGGA
7081 CTTATAAAAA TAAGGTAGTG AAGTAGCAGC AGGGCTTAAA TAAAAACCCA TGTCTGTACC
7141 AACCACAGAG TCACCCATCC AGGTTAAAAT AACCAGAGAA ACAGAAGATA TTCCTACTAC
7201 AGAGAATTCC GGGTGTGCAG CCACAGTGCA AATCCTTTTT ATTTTTATTT TTGAGATGCA
7261 GTCTCGCTCT GTCATCCAGG CTGAAGTGCA GTGGCACGAT CATGTCTCGC TGCAACCTCT

7321 GCCTCCCAGG CTCAAGCGAT CCTCCCACCT CAGCCATCTG AGTAGCTGGG ACCACAGGCC
7381 ACACACCACA CCCAGCTAAT TTCTCGTATC TTTTTGTAGA GACAGAGTTC TGCTATGTTG
7441 CCCAGGCTCA GGCTGGTCTT GATCTCAAGC AATTGGCTTG CCTCAGCCTC CTAAAATATT
7501 GGGATTACAG GCATGAGCCA CCGCGCCAGC CATGCAAATC CTTAATTATC AAACAGATAA
7561 AATAGGGAAG TTAAAATTCA TATACACAAG GGTTAACCAC TTGCCACAGG CATTTTTTTT
7621 TTTTTTTTGA GACGGAATCT CGCTCTGTTG CCCAGGCTGG AGTGCAGTGG CGCCATCTCG
7681 CCTCACTGCA ACCTCCGCTT CCTGGGTTCA AGCTATTCTT CTGCCTCAGC CTACCGAGTA
7741 GCTGGGACTA CAGGCACGTG CCACCACACC TGGCTAATTT TTTTATTTTT AGTAGAGATG
7801 GGGTTTCACC ATATTGGCCA GGCTGGTCTT GAACTCCTGA CCTAGTGATC CATCCGCCTC
7861 AGCCTCCCAA AGTGCTGGGA TTGCAGGCAT GAGCCACCGC GCCTGGCCTT TTTTTTTTTT
7921 TTTTGAGACG GAGTTTTGCT CTTGTTGCCC AGGCTAGAGT GCAGTGGCGC AGTCTCGGCT
7981 CACTGTAACC TCCACCTCCT GAGTTCAAGC AATTCTCCTG CCTCAGCCTC TCAAATAGCT
8041 GGGATTACAG GCGTGAGCCA CCCCACCTGG CTAATTTTGT AATTTTTTTT TTAGTAGAGA
8101 TGGGGTTTCA CCTGTTGATC AGGCTGGTCT CAAACTCCTG ACCTCAAGTG ATCCACCCAC
8161 CTCGGCCTCC CAAAGTGCTG GGATTACAAG CATAAGCCAC CGTGCCTGGT CAATTTTGAT
8221 CTTTTTTAAA GAGACAGGGG TCTTGCTATG TTGCCCAGAC TAGTCTTGAA CTCCTGGCCT
8281 CAAGTGATCC TCTCACCTCG GCCTCCCAAA GTATTGGGAT TACAGGTCTG AGCCGCTGCA
8341 CCCAGCCCCC AACAGGCATC TTTGGACTTT TGAGTACTGG CTTTAATTTA CAAAAATTCC
8401 ACTGAGAGCA CCTAAGTTTG CCAGGCTCCA ACATTTCTGC AGGGGCTGTT TTCTTTGCTG
8461 AAGGATCTGC ACCTGTGTTC TGTTATGGTT GCCTCTTCTG TTGCAGGTGC TTGAATGTCA
8521 TTTTGTGGTT GGGATTCTGG GCTGTGCAGC TGAATGTCTG TCTGTCACGA ATCTACCTTG
8581 CTGCTCATTT TCCTCATCAA GTTGTTGCTG GAGTCCTGTC AGGTATGGGC TGATCTGACT
8641 CCCTTCCTTC TCCCCCAAAC CCCATTCCGT TTCTCTCCCT AATCAGGACA AAATCCCAGC
8701 ATTCCAGCCA CATCCTGTGT GTAATCAGTA CTGTTAGCAT TTCTGTGGGT TGAAAGTCAA
8761 GAATGAGCAA CTTGAAATGA TTAATTTCTA TAAGAGTGCC CAGATCTATA GAATGAATTG
8821 TGTAGAAGTT ACCATACATC AAATTAACGC ACCAAATTGA ATTAGCTTGA AATCTCAGAG
8881 CTTTTTACAA TCTTTATTTC TTACTGGTCT TCAACAGGCC CTAATTTACT TTTCAGGGAA
8941 TCTGCCAAAT TTAACAAATT AACACGATGT CCTAGGAAAG CTGTTCATTT AAATACATTC
9001 ATTTGCAAAC CTAATAGATA ACTGCAGTTG ATCTCTTTTA TAGGTTCAGA GTTTTGAATA
9061 TGTTTTTTTT TGTTTTTTTT TTTTGAGATG GAGTCTCGCT CTGTGACCCA GGCTAGAGTG
9121 CAGTGGTGCG ATCTCGGCTC ACTGCAAGCT CCACCTCCTG GGTTCACGCC ATTCTCCTGC
9181 CTCAGCCTCT CCGAGTAGCT GGGACTACAG GCGCCCGCCA CCATGCCCGG CTAATTTTTT
9241 GTATTTTTAG CAGAGACGGG GTTTCACCGT GGTCTTGATC TCCTGACCTC GTGATCCGCC
9301 CGCCTCGGCC TCCCAAAGCG CTGGGATTAC AAGGGTGAGC CACCGCACCC TGCCTGAATA
9361 TGTGTTTTCT TAGATCCAAT TAACAAGGGT AAGACAAGAT TTAAGTTAAG CATAAGAAAG
9421 ATTTTGTGGG AGGCACTGGA ATATAAGACC TTAACAAAAC TGTGGAATTT CTCCCCTGGA
9481 GATTTGTAAG AACGGAACAT AGCAGCATTC AAAGAAGAAT GTTGAGAACA AGGGAGATAA
9541 TGGTTTCATG GTAATCACAA AAGTAACACA GCATTTAGTA CTGGGTTCCA TGTTTGAGGA
9601 AGAACCTGGA AGCCATATCA CATGAAAAAC CTGGGAATGT TTAGGTTAGA GAGAATAACT
9661 GTGTTCAAAT GTGTGACAGA GGGACTAGAT TCATCACTTA CTAACTCCTG CAGAAAGAAC

```
9721 TGAGAAAAAT AGACAGTATT AGAGGGGGAC CAGTTTCACA CAGACAAGGA AGAACTATTC
9781 AGCAATCAAT TCCGTTCAAA GATAAAATGG ACTGTTATAG TGGGGGTGAG CTCCCTACCT
9841 CTGAGGGTAT TTCAAGTAGA GATAGGAGGA CCTCCTGGTA GGAAATTTGC ATACGGTGGG
9901 AGATTGTACG TGATATGGCA CCTCCATCTG AAAGAGTCTA TATTGAGGGC AGGCTGGAGT
9961 CACACATGGG AATAAGCCAG GCGACCCTCC CATCTGCCAT CTGTGATTTA ATTCCACAGT
10021 CGCAGAACGG ATGGCATGTC ACCCACTCCT CCAAACCCAC CTCTAGCAAA GGTCCCAAAT
10081 CCTTCCTATC TCTCACAGTC ATGCTTTCTT CCACTCAGGC ATTGCTGTTA CAGAAACTTT
10141 CAGCCACATC CACAGCATCT ATAATGCCAG CCTCAAGAAA TATTTTCTCA TTACCTTCTT
10201 CCTGTTCAGC TTCGCCATCG GATTTTATCT GCTGCTCAAG GGACTGGGTG TAGACCTCCT
10261 GTGGACTCTG GAGAAAGCCC AGAGGTGGTG CGAGCAGCCA GAATGGGTCC ACATTGACAC
10321 CACACCCTTT GCCAGCCTCC TCAAGAACCT GGGCACGCTC TTTGGCCTGG GGCTGGCTCT
10381 CAACTCCAGC ATGTACAGGG AGAGCTGCAA GGGGAAACTC AGCAAGTGGC TCCCATTCCG
10441 CCTCAGCTCT ATTGTAGCCT CCCTCGTCCT CCTGCACGTC TTTGACTCCT TGAAACCCCC
10501 ATCCCAAGTC GAGCTGGTCT TCTACGTCTT GTCCTTCTGC AAGAGTGCGG TAGTGCCCCT
10561 GGCATCCGTC AGTGTCATCC CCTACTGCCT CGCCCAGGTC CTGGGCCAGC CGCACAAGAA
10621 GTCGTTGTAA GAGATGTGGA GTCTTCGGTG TTTAAAGTCA ACAACCATGC CAGGGATTGA
10681 GGAGGACTAC TATTTGAAGC AATGGGCACT GGTATTTGGA GCAAGTGACA TGCCATCCAT
10741 TCTGCCGTCG TGGAATTAAA TCACGGATGG CAGATTGGAG GGTCGCCTGG CTTATTCCCA
10801 TGTGTGACTC CAGCCTGCCC TCAGCACAGA CTCTTTCAGA TGGAGGTGCC ATATCACGTA
10861 CACCATATGC AAGTTTCCCG CCAGGAGGTC CTCCTCTCTC TACTTGAATA CTCTCACAAG
10921 TAGGGAGCTC ACTCCCACTG GAACAGCCCA TTTTATCTTT GAATGGTCTT CTGCCAGCCC
10981 ATTTTGAGGC CAGAGGTGCT GTCAGCTCAG GTGGTCCTCT TTTACAATCC TAATCATATT
11041 GGGTAATGTT TTTGAAAAGC TAATGAAGCT ATTGAGAAAG ACCTGTTGCT AGAAGTTGGG
11101 TTGTTCTGGA TTTTCCCCTG AAGACTTACT TATTCTTCCG TCACATATAC AAAAGCAAGA
11161 CTTCCAGGTA GGGCCAGCTC ACAAGCCCAG GCTGGAGATC CTAACTGAGA ATTTTCTACC
11221 TGTGTTCATT CTTACCGAGA AAAGGAGAAA GGAGCTCTGA ATCTGATAGG AAAAGAAGGC
11281 TGCCTAAGGA GGAGTTTTTA GTATGTGGCG TATCATGCAA GTGCTATGCC AAGCCATGTC
11341 TAAATGGCTT TAATTATATA GTAATGCACT CTCAGTAATG GGGGACCAGC TTAAGTATAA
11401 TTAATAGATG GTTAGTGGGG TAATTCTGCT TCTAGTATTT TTTTTACTGT GCATACATGT
11461 TCATCGTATT TCCTTGGATT TCTGAATGGC TGCAGTGACC CAGATATTGC ACTAGGTCAA
11521 AACATTCAGG TATAGCTGAC ATCTCCTCTA TCACATTACA TCATCCTCCT TATAAGCCCA
11581 GCTCTGCTTT TTCCAGATTC TTCCACTGGC TCCACATCCA CCCCACTGGA TCTTCAGAAG
11641 GCTAGAGGGC GACTCTGGTG GTGCTTTTGT ATGTTTCAAT TAGGCTCTGA AATCTTGGGC
11701 AAAATGACAA GGGGAGGGCC AGGATTCCTC TCTCAGGTCA CTCCAGTGTT ACTTTTAATT
11761 CCTAGAGGGT AAATATGACT CCTTTCTCTA TCCCAAGCCA ACCAAGAGCA CATTCTTAAA
11821 GGAAAAGTCA ACATCTTCTC TCTTTTTTTT TTTTTTTGAG ACAGGGTCTC ACTATGTTGC
11881 CCAGGCTGCT CTTGAATTCC TGGGCTCAAG CAGTCCTCCC ACCCTACCAC AGCGTCCCGC
11941 GTAGCTGGGA CTACAGGTGC AAGCCACTAT GTCCAGCTAG CCAACTCCTC CTTGCCTGCT
12001 TTTCTTTTTT TTTCTTTTTT TGAGACGGCG CACCTATCAC CCAGGCTGGA GTGGAGTGGC
12061 ACGATCTTGG CTCACTGCAA CCTCTTCCTC CTGGTTCAAG CGATTCTCAT GTCTCAGCCT
```

12121 CCTCAGTAGC TAGGACTACC GGCGTGCACC ACCATGCCAG GCTAATTTTT ATATTTTTAG
12181 AATTTTAGAA GAGATGGGAT TTCATCATGT TGGCCAGGCT GGTCTCGAAC TCCTGACCTC
12241 AAGTGATCCA CCTGCCTTGG CCTCCCAAGG TGCTAGGATT ACAGGCATGA GCCACCGCAC
12301 CGGGCCCTCC TTGCCTGTTT TTCAATCTCA TCTGATATGC AGAGTATTTC TGCCCCACCC
12361 ACCTACCCCC CAAAAAAAGC TGAAGCCTAT TTATTTGAAA GTCCTTGTTT TTGCTACTAA
12421 TTATATAGTA TACCATACAT TATCATTCAA AACAACCATC CTGCTCATAA CATCTTTGAA
12481 AAGAAAAATA TATATGTGCA GTATTTTATT AAAGCAACAT TTTATTTAAG AATAAAGTCT
12541 TGTTAATTAC TATATTTTAG ATGCAATGTG ATCTGAAGTT TCTAATTCTG GCCCAACTAA
12601 ATTTCTAGCT CTGTTTCCCT AAACAAATAA TTTGGTTTCT CTGTGCCTGC ATTTTCCCTT
12661 TGGAGAAGAA AAGTGCTCTC TCTTGAGTTG ACCGAGAGTC CCATTAGGGA TAGGGAGACT
12721 TAAATGCATC CACAGGGGCA CAGGCAGAGT TGAGCACATA AACGGAGGCC CAAAATCAGC
12781 ATAGAACCAG AAAGATTCAG AGTTGGCCAA GAATGAACAT TGGCTACCAG ACCACAAGTC
12841 AGCATGAGTT GCTCTATGGC ATCAAATTGC AACTTGAGAG TAGATGGGCA GGGTCACTAT
12901 CAAATTAAGC AATCAGGGCA CACAAGTTGC AGTAACACAA CAAGACTAGG CCAGCTCTGG
12961 AATCCAGTAA CTCAGTGTCA GCAAGGTTTT GGGTTATAGT TCAAGAAAGT CTAAACAGAG
13021 CCAGTCACAG CACCAAGGAA TGCTCAAGGG AGCTATTGCA GGTTTCTCTG CTAAGAGATT
13081 TATTTCATCC TGGGTGCAGG GTTCGACCTC CAAAGGCCTC AAATCATCAC CGTATCAATG
13141 GATTTCCTGA GGGTAAGCTC CGCTATTTCA CACCTGAACT CCGGAGTCTG TATATTCAGG
13201 GAAGATTGCA TTCTCCTACT GGATTTGGGC TCTCAGAGGG CGTTGTGGGA ACCAGGCCCC
13261 TCACAGAATC AAATGGTCCC AACCAGGGAG AAAGAAAATA GTCTTTTTTT TTTTTTAAT
13321 AGAGATGGGG GTCTCACTAT GCTGCCCAGG CTGGTCTTGA ACTCCTGGGT TCAAGTGATC
13381 CTCCTGCCTC AGCCTCCCAA AGTGCTGGGA TTACAGTGTG AGCCACTGCG CTTGGCCAGA
13441 AATGGTTTTG ATCTGTCTGA ACTGAACCCT ACTGCTTAGG CATAGCCCCA TCCTTGATAA
13501 TCTATTTGCT CCCAAGGACC AAGTCCAAGA TCCTTACAAG AAAGGTCTGC CAGAAAGTAA
13561 ATACTGCCCC CACTCCCTGA AGTTTATGAG GTTGATAAGA AAACATAACA GATAAAGTTT
13621 ATTGAGTGCT AACTTTA (SEQ ID NO: 380).

By "guide polynucleotide" is meant a polynucleotide or polynucleotide complex which is specific for a target sequence and can form a complex with a polynucleotide programmable nucleotide binding domain protein (*e.g.*, Cas9 or Cpf1). In an embodiment, the guide polynucleotide is a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule.

By "heterodimer" is meant a fusion protein comprising two domains, such as a wild type TadA domain and a variant of TadA domain or two variant TadA domains.

By "heterologous," or "exogenous" is meant a polynucleotide or polypeptide that 1) has been experimentally incorporated to a polynucleotide or polypeptide sequence to which the polynucleotide or polypeptide is not normally found in nature; or 2) has been experimentally placed into a cell that does not normally comprise the polynucleotide or polypeptide. In some embodiments, "heterologous" means that a polynucleotide or

polypeptide has been experimentally placed into a non-native context. In some embodiments, a heterologous polynucleotide or polypeptide is derived from a first species or host organism, and is incorporated into a polynucleotide or polypeptide derived from a second species or host organism. In some embodiments, the first species or host organism is different from the second species or host organism. In some embodiments the heterologous polynucleotide is DNA. In some embodiments the heterologous polynucleotide is RNA.

"Hybridization" means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleobases. For example, adenine and thymine are complementary nucleobases that pair through the formation of hydrogen bonds.

By "increases" is meant a positive alteration of at least 10%, 25%, 50%, 75%, or 100%.

The term "inhibitor of base repair" or "IBR" refers to a protein that is capable in inhibiting the activity of a nucleic acid repair enzyme, for example a base excision repair (BER) enzyme. In some embodiments, the IBR is an inhibitor of inosine base excision repair. Exemplary inhibitors of base repair include inhibitors of APE1, Endo III, Endo IV, Endo V, Endo VIII, Fpg, hOGGl, hNEILl, T7 Endol, T4PDG, UDG, hSMUGl, and hAAG. In some embodiments, the IBR is an inhibitor of Endo V or hAAG. In some embodiments, the IBR is a catalytically inactive EndoV or a catalytically inactive hAAG. In some embodiments, the base repair inhibitor is an inhibitor of Endo V or hAAG. In some embodiments, the base repair inhibitor is a catalytically inactive EndoV or a catalytically inactive hAAG.

In some embodiments, the base repair inhibitor is uracil glycosylase inhibitor (UGI). UGI refers to a protein that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a fragment of a wild-type UGI. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. In some embodiments, the base repair inhibitor is an inhibitor of inosine base excision repair. In some embodiments, the base repair inhibitor is a "catalytically inactive inosine specific nuclease" or "dead inosine specific nuclease. Without wishing to be bound by any particular theory, catalytically inactive inosine glycosylases (*e.g.*, alkyl adenine glycosylase (AAG)) can bind inosine, but cannot create an abasic site or remove the inosine, thereby sterically blocking the newly formed inosine moiety from DNA damage/repair mechanisms. In some embodiments,

the catalytically inactive inosine specific nuclease can be capable of binding an inosine in a nucleic acid but does not cleave the nucleic acid. Non-limiting exemplary catalytically inactive inosine specific nucleases include catalytically inactive alkyl adenosine glycosylase (AAG nuclease), for example, from a human, and catalytically inactive endonuclease V (EndoV nuclease), for example, from *E. coli*. In some embodiments, the catalytically inactive AAG nuclease comprises an E125Q mutation or a corresponding mutation in another AAG nuclease.

An "intein" is a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing.

The terms “isolated,” “purified,” or “biologically pure” refer to material that is free to varying degrees from components which normally accompany it as found in its native state. “Isolate” denotes a degree of separation from original source or surroundings. “Purify” denotes a degree of separation that is higher than isolation. A “purified” or “biologically pure” protein is sufficiently free of other materials such that any impurities do not materially affect the biological properties of the protein or cause other adverse consequences. That is, a nucleic acid or peptide as described herein is purified if it is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Purity and homogeneity are typically determined using analytical chemistry techniques, for example, polyacrylamide gel electrophoresis or high-performance liquid chromatography. The term “purified” can denote that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. For a protein that can be subjected to modifications, for example, phosphorylation or glycosylation, different modifications may give rise to different isolated proteins, which can be separately purified.

By “isolated polynucleotide” is meant a nucleic acid (*e.g.*, a DNA) that is free of the genes which, in the naturally-occurring genome of the organism from which the nucleic acid molecule as described is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. In addition, the term includes an RNA molecule that is transcribed from a DNA molecule, as well as a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

By an "isolated polypeptide" is meant a polypeptide as described that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation comprises at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a polypeptide as described herein. An isolated polypeptide as described herein may be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding such a polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

The term "linker", as used herein, refers to a molecule that links two moieties. In one embodiment, the term "linker" refers to a covalent linker (*e.g.*, covalent bond) or a non-covalent linker.

By "marker" is meant any protein or polynucleotide having an alteration in expression level or activity that is associated with a disease or disorder, such as, GSD1a and/or symptoms thereof..

The term "mutation," as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

The term "non-conservative mutations" involve amino acid substitutions between different groups, for example, lysine for tryptophan, or phenylalanine for serine, etc. In this case, it is preferable for the non-conservative amino acid substitution to not interfere with, or inhibit the biological activity of, the functional variant. The non-conservative amino acid substitution can enhance the biological activity of the functional variant, such that the biological activity of the functional variant is increased as compared to the wild-type protein.

The term "nuclear localization sequence," "nuclear localization signal," or "NLS" refers to an amino acid sequence that promotes import of a protein into the cell nucleus. Nuclear localization sequences are known in the art and described, for example, in Plank *et*

al., International PCT application, PCT/EP2000/011690, filed November 23, 2000, published as WO/2001/038547 on May 31, 2001, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In other embodiments, the NLS is an optimized NLS described, for example, by Koblan *et al.*, Nature Biotech. 2018 doi:10.1038/nbt.4172. In some embodiments, an NLS comprises the amino acid sequence In some embodiments, an NLS comprises the amino acid sequence KRTADGSEFESPKKKRKV (SEQ ID NO: 190), KRPAATKKAGQAKKKK (SEQ ID NO: 191), KKTELQTTNAENKTKKL (SEQ ID NO: 192), KRGINDRNFWRGENGRKTR (SEQ ID NO: 193), RKSGKIAAIVVKRPRK (SEQ ID NO: 194), PKKKRKV (SEQ ID NO: 195), or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 196).

The term "nucleobase," "nitrogenous base," or "base," used interchangeably herein, refers to a nitrogen-containing biological compound that forms a nucleoside, which in turn is a component of a nucleotide. The ability of nucleobases to form base pairs and to stack one upon another leads directly to long-chain helical structures such as ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). Five nucleobases – adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U) – are called primary or canonical. Adenine and guanine are derived from purine, and cytosine, uracil, and thymine are derived from pyrimidine. DNA and RNA can also contain other (non-primary) bases that are modified. Non-limiting exemplary modified nucleobases can include hypoxanthine, xanthine, 7-methylguanine, 5,6-dihydrouracil, 5-methylcytosine (m5C), and 5-hydromethylcytosine. Hypoxanthine and xanthine can be created through mutagen presence, both of them through deamination (replacement of the amine group with a carbonyl group). Hypoxanthine can be modified from adenine. Xanthine can be modified from guanine. Uracil can result from deamination of cytosine. A "nucleoside" consists of a nucleobase and a five carbon sugar (either ribose or deoxyribose). Examples of a nucleoside include adenosine, guanosine, uridine, cytidine, 5-methyluridine (m5U), deoxyadenosine, deoxyguanosine, thymidine, deoxyuridine, and deoxycytidine. Examples of a nucleoside with a modified nucleobase includes inosine (I), xanthosine (X), 7-methylguanosine (m7G), dihydrouridine (D), 5-methylcytidine (m5C), and pseudouridine (Ψ). A "nucleotide" consists of a nucleobase, a five carbon sugar (either ribose or deoxyribose), and at least one phosphate group.

The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules

comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *e.g.*, analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc*. Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, 2′-fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5′-*N*-phosphoramidite linkages). In some embodiments, a polynucleotide described herein comprises one or more of the following modifications: 2′-*O*-methyl (2′-OMe),

phosphorothioate (PS), 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), and constrained ethyl (S-cEt).

The term "nucleic acid programmable DNA binding protein" or "napDNAbp" may be used interchangeably with "polynucleotide programmable nucleotide binding domain" to refer to a protein that associates with a nucleic acid (*e.g.*, DNA or RNA), such as a guide nucleic acid or guide polynucleotide (*e.g.*, gRNA), that guides the napDNAbp to a specific nucleic acid sequence. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a Cas9 protein. A Cas9 protein can associate with a guide RNA that guides the Cas9 protein to a specific DNA sequence that is complementary to the guide RNA. In some embodiments, the napDNAbp is a Cas9 domain, for example a nuclease active Cas9, a Cas9 nickase (nCas9), or a nuclease inactive Cas9 (dCas9). Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ (Cas12j/Casphi). Non-limiting examples of Cas enzymes include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (also known as Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl (*e.g.*, SEQ ID NO: 232), Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Cas12j/CasΦ, Cpf1, Csy1, Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, homologues thereof, or modified or engineered versions thereof. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, although they may not be specifically listed in this disclosure. See, *e.g.*, Makarova *et al.* "Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?" *CRISPR J.* 2018 Oct;1:325-336. doi: 10.1089/crispr.2018.0033; Yan *et al.*, "Functionally diverse type V CRISPR-Cas systems" *Science*. 2019 Jan 4;363(6422):88-91. doi: 10.1126/science.aav7271, the entire contents of each are hereby incorporated by reference. Exemplary nucleic acid

programmable DNA binding proteins and nucleic acid sequences encoding nucleic acid programmable DNA binding proteins are provided in the Sequence Listing as SEQ ID NOs: 197-230.

The terms "nucleobase editing domain" or "nucleobase editing protein," as used herein, refers to a protein or enzyme that can catalyze a nucleobase modification in RNA or DNA, such as cytosine (or cytidine) to uracil (or uridine) or thymine (or thymidine), and adenine (or adenosine) to hypoxanthine (or inosine) deaminations, as well as non-templated nucleotide additions and insertions. In some embodiments, the nucleobase editing domain is a deaminase domain (*e.g.*, an adenine deaminase or an adenosine deaminase; or a cytidine deaminase or a cytosine deaminase).

As used herein, "obtaining" as in "obtaining an agent" includes synthesizing, purchasing, or otherwise acquiring the agent.

A "patient" or "subject" as used herein refers to a mammalian subject or individual diagnosed with, at risk of having or developing, or suspected of having or developing a disease or a disorder. In some embodiments, the term "patient" refers to a mammalian subject with a higher than average likelihood of developing a disease or a disorder. Exemplary patients can be humans, non-human primates, cats, dogs, pigs, cattle, cats, horses, camels, llamas, goats, sheep, rodents (*e.g.*, mice, rabbits, rats, or guinea pigs) and other mammalians that can benefit from the therapies disclosed herein. Exemplary human patients can be male and/or female.

"Patient in need thereof" or "subject in need thereof" is referred to herein as a patient diagnosed with, at risk or having, predetermined to have, or suspected of having a disease or disorder, for instance, but not restricted to Glycogen Storage Disease Type 1 (GSD1 or Von Gierke Disease).

The terms "pathogenic mutation," "pathogenic variant," "disease casing mutation," "disease causing variant," "deleterious mutation," or "predisposing mutation" refers to a genetic alteration or mutation that increases an individual's susceptibility or predisposition to a certain disease or disorder. In some embodiments, the pathogenic mutation comprises at least one wild-type amino acid substituted by at least one pathogenic amino acid in a protein encoded by a gene.

The term "pharmaceutically-acceptable carrier" means a pharmaceutically-acceptable material, composition, or vehicle, such as a liquid or solid filler, diluent, excipient, manufacturing aid (*e.g.*, lubricant, talc magnesium, calcium or zinc stearate, or steric acid), or

solvent encapsulating material, involved in carrying or transporting the compound from one site (*e.g.*, the delivery site) of the body, to another site (*e.g.*, organ, tissue or portion of the body). A pharmaceutically acceptable carrier is "acceptable" in the sense of being compatible with the other ingredients of the formulation and not injurious to the tissue of the subject (*e.g.*, physiologically compatible, sterile, physiologic pH, etc.). The terms such as "excipient," "carrier," "pharmaceutically acceptable carrier," "vehicle," or the like are used interchangeably herein.

The term "pharmaceutical composition" means a composition formulated for pharmaceutical use. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier. In some embodiments, the pharmaceutical composition comprises additional agents (*e.g.*, for specific delivery, increasing half-life, or other therapeutic compounds).

The terms "protein", "peptide", "polypeptide", and their grammatical equivalents are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. A protein, peptide, or polypeptide can be naturally occurring, recombinant, or synthetic, or any combination thereof.

The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins.

By "promoter" is meant an array of nucleic acid control sequences, which direct transcription of a nucleic acid. A promoter includes necessary nucleic acid sequences near the start site of transcription. A promoter also optionally includes distal enhancer or repressor sequence elements. A "constitutive promoter" is a promoter that is continuously active and is not subject to regulation by external signals or molecules. In contrast, the activity of an "inducible promoter" is regulated by an external signal or molecule (for example, a transcription factor). By way of example, a promoter may be a CMV promoter.

The term "recombinant" as used herein in the context of proteins or nucleic acids refers to proteins or nucleic acids that do not occur in nature, but are the product of human engineering. For example, in some embodiments, a recombinant protein or nucleic acid molecule comprises an amino acid or nucleotide sequence that comprises at least one, at least two, at least three, at least four, at least five, at least six, or at least seven mutations as compared to any naturally occurring sequence.

By "reduces" is meant a negative alteration of at least 10%, 25%, 50%, 75%, or 100%.

By "reference" is meant a standard or control condition. In one embodiment, the reference is a wild-type or healthy cell. For example, in some embodiments the reference is a cell of a subject not inflicted with GSD1a. In some embodiments, the reference is a cell with normal or wild-type glucose-6-phosphatase activity. In other embodiments and without limitation, a reference is an untreated cell that is not subjected to a test condition, or is subjected to placebo or normal saline, medium, buffer, and/or a control vector that does not harbor a polynucleotide of interest.

A "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset of or the entirety of a specified sequence; for example, a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. For polypeptides, the length of the reference polypeptide sequence will generally be at least about 16 amino acids, at least about 20 amino acids, at least about 25 amino acids, about 35 amino acids, about 50 amino acids, or about 100 amino acids. For nucleic acids, the length of the reference nucleic acid sequence will generally be at least about 50 nucleotides, at least about 60 nucleotides, at least about 75 nucleotides, about 100 nucleotides or about 300 nucleotides or any integer thereabout or therebetween. In some embodiments, a reference sequence is a wild-type sequence of a protein of interest. In other embodiments, a reference sequence is a polynucleotide sequence encoding a wild-type protein.

The term "RNA-programmable nuclease," and "RNA-guided nuclease" are used with (*e.g.*, binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csnl) from *Streptococcus pyogenes*.

The term "single nucleotide polymorphism (SNP)" is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (*e.g.*, > 1%).

By "specifically binds" is meant a nucleic acid molecule, polypeptide, or complex thereof (*e.g.*, a nucleic acid programmable DNA binding protein, a guide nucleic acid), compound, or molecule that recognizes and binds a polypeptide and/or nucleic acid molecule

as described herein, but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample.

By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence. In one embodiment, a reference sequence is a wild-type amino acid or nucleic acid sequence. In another embodiment, a reference sequence is any one of the amino acid or nucleic acid sequences described herein. In one embodiment, such a sequence is at least 60%, 80%, 85%, 90%, 95% or even 99% identical at the amino acid level or nucleic acid level to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

COBALT is used, for example, with the following parameters:

a) alignment parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1,

b) CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and

c) Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

EMBOSS Needle is used, for example, with the following parameters:

a) Matrix: BLOSUM62;

b) GAP OPEN: 10;

c) GAP EXTEND: 0.5;

d) OUTPUT FORMAT: pair;

e) END GAP PENALTY: false;

f) END GAP OPEN: 10; and

g) END GAP EXTEND: 0.5.

Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. Nucleic acid molecules useful in the methods of the invention include any nucleic acid molecule that encodes a polypeptide of the invention or a fragment thereof. Such nucleic acid molecules need not be 100% identical with an endogenous nucleic acid sequence, but will typically exhibit substantial identity. Polynucleotides having "substantial identity" to an endogenous sequence are typically capable of hybridizing with at least one strand of a double-stranded nucleic acid molecule. By "hybridize" is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (*e.g.*, a gene described herein), or portions thereof, under various conditions of stringency. (See, *e.g.*, Wahl, G. M. and S. L. Berger (1987) Methods Enzymol. 152:399; Kimmel, A. R. (1987) Methods Enzymol. 152:507).

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and more preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, *e.g.*, formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and more preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, *e.g.*, sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred: embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 μg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate,

1% SDS, 50% formamide, and 200 μg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and even more preferably of at least about 68° C. In an embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In another embodiment, wash steps will occur at 42 C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel *et al.* (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

By “split” is meant divided into two or more fragments.

A "split Cas9 protein" or "split Cas9" refers to a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a “reconstituted” Cas9 protein.

By “subject” is meant a mammal, including, but not limited to, a human or non-human mammal, such as a bovine, equine, canine, ovine, or feline. Subjects include livestock, domesticated animals raised to produce labor and to provide commodities, such as food, including without limitation, cattle, goats, chickens, horses, pigs, rabbits, and sheep.

By “substantially identical” is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of

the nucleic acid sequences described herein). In one embodiment, such a sequence is at least 60%, 80% or 85%, 90%, 95% or even 99% identical at the amino acid level or nucleic acid to the sequence used for comparison.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, COBALT, EMBOSS Needle, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

COBALT is used, for example, with the following parameters:

a) alignment parameters: Gap penalties-11,-1 and End-Gap penalties-5,-1,

b) CDD Parameters: Use RPS BLAST on; Blast E-value 0.003; Find Conserved columns and Recompute on, and

c) Query Clustering Parameters: Use query clusters on; Word Size 4; Max cluster distance 0.8; Alphabet Regular.

EMBOSS Needle is used, for example, with the following parameters:

a) Matrix: BLOSUM62;

b) GAP OPEN: 10;

c) GAP EXTEND: 0.5;

d) OUTPUT FORMAT: pair;

e) END GAP PENALTY: false;

f) END GAP OPEN: 10; and

g) END GAP EXTEND: 0.5.

The term “target site” refers to a sequence within a nucleic acid molecule that is modified by a nucleobase editor. In one embodiment, the target site is deaminated by a deaminase or a fusion protein comprising a deaminase (*e.g.*, adenine deaminase).

As used herein “transduction” means to transfer a gene or genetic material to a cell via a viral vector.

"Transformation," as used herein refers to the process of introducing a genetic change in a cell produced by the introduction of exogenous nucleic acid.

"Transfection" refers to the transfer of a gene or genetical material to a cell via a chemical or physical means.

By "translocation" is meant the rearrangement of nucleic acid segments between non-homologous chromosomes.

As used herein, the terms "treat," treating," "treatment," and the like refer to reducing or ameliorating a disorder and/or symptom(s) associated therewith or obtaining a desired pharmacologic and/or physiologic effect. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated. In some embodiments, the effect is therapeutic, *i.e.*, without limitation, the effect partially or completely reduces, diminishes, abrogates, abates, alleviates, decreases the intensity of, or cures a disease and/or adverse symptom attributable to the disease. In some embodiments, the effect is preventative, *i.e.*, the effect protects or prevents an occurrence or reoccurrence of a disease or condition. To this end, the presently disclosed methods comprise administering a therapeutically effective amount of a compositions as described herein.

By "uracil glycosylase inhibitor" or "UGI" is meant an agent that inhibits the uracil-excision repair system. In one embodiment, the agent is a protein or fragment thereof that binds a host uracil-DNA glycosylase and prevents removal of uracil residues from DNA. In an embodiment, a UGI is a protein, a fragment thereof, or a domain that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a modified version thereof. In some embodiments, a UGI domain comprises a fragment of the exemplary amino acid sequence set forth below. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the exemplary UGI sequence provided below. In some embodiments, a UGI comprises an amino acid sequence that is homologous to the exemplary UGI amino acid sequence or fragment thereof, as set forth below. In some embodiments, the UGI, or a portion thereof, is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.9%, or 100% identical to a wild-

type UGI or a UGI sequence, or portion thereof, as set forth below. An exemplary UGI comprises an amino acid sequence as follows:

>splP14739IUNGI_BPPB2 Uracil-DNA glycosylase inhibitor

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDA

PEYKPWALVIQDSNGENKIKML (SEQ ID NO: 231).

The term "vector" refers to a means of introducing a nucleic acid sequence into a cell, resulting in a transformed cell. Vectors include plasmids, transposons, phages, viruses, liposomes, and episome. "Expression vectors" are nucleic acid sequences comprising the nucleotide sequence to be expressed in the recipient cell. Expression vectors may include additional nucleic acid sequences to promote and/or facilitate the expression of the of the introduced sequence such as start, stop, enhancer, promoter, and secretion sequences.

The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable or aspect herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

Any compositions or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

DNA editing has emerged as a viable means to modify disease states by correcting pathogenic mutations at the genetic level. Until recently, all DNA editing platforms have functioned by inducing a DNA double strand break (DSB) at a specified genomic site and relying on endogenous DNA repair pathways to determine the product outcome in a semi-stochastic manner, resulting in complex populations of genetic products. Though precise, user-defined repair outcomes can be achieved through the homology directed repair (HDR) pathway, a number of challenges have prevented high efficiency repair using HDR in therapeutically-relevant cell types. In practice, this pathway is inefficient relative to the competing, error-prone non-homologous end joining pathway. Further, HDR is tightly restricted to the G1 and S phases of the cell cycle, preventing precise repair of DSBs in post-mitotic cells. As a result, it has proven difficult or impossible to alter genomic sequences in a user-defined, programmable manner with high efficiencies in these populations.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 provides a schematic depicting a G6PC nucleotide target sequence (ATTCTCTTTGGACAGTGTCCATACTGGTGG; SEQ ID NO 399) and corresponding amino

acid sequence (ILFGQCPYWW; SEQ ID NO: 401) indicating bystander and on target A > G bases for correction of the GSD1a R83C mutation.

FIGs. 2A and 2B depict *in vivo* correction of GSD1a mutations in liver extracts of transgenic mouse models heterozygous for huG6PC-R83C. **FIG. 2A** is a schematic depicting *in vivo* workflow. Lipid nanoparticles (LNP) carrying base editor mRNA and gRNA were dosed via IV injection in transgenic mice heterozygous for huG6PC (huR83C HET), harboring the R83C mutation. **FIG. 2B** is a bar graph depicting A to G base editing efficiency of the GSD1a R83C mutation using MSP828 comparing on-target to bystander editing.

FIG. 3 is a bar graph depicting correction of the GSD1a R83C mutation in a transgenic mouse model heterozygous for huG6PC, harboring the R83C mutation, using TadA adenosine deaminase variants MSP605, MSP824, MSP825, MSP680, MSP828, and MSP820. *In vitro* screens were run to select desirable base-editors for R83C correction. LNP co-formulations of gRNA and representative base-editors were dosed (at a sub-saturating dose of 1mpk), *in vivo*, in transgenic mice heterozygous for huG6PC-R83C. The base-editing potency of the variants for the R83C correction in livers of the LNP-treated, huG6PC-R83C heterozygote, transgenic animals are shown in FIG. 3. Variant MSP828 yielded a high level of on-target activity under these conditions. A to G base editing efficiency is shown for on-target and bystander editing.

FIG. 4 shows schematics depicting normal and loss-of-function *g6pc* function and related outcomes. GSD-Ia (or GSD1a herein) is an autosomal recessive disorder caused by mutations in the *g6pc* gene. R83C, located in the active site of the enzyme, is the most prevalent pathogenic mutation identified in Caucasian GSD-Ia patients and is associated with inactivation of G6Pase. A loss of G6Pase function can result in life-threatening hypoglycemia, seizures and even death. To mitigate hypoglycemia, patients must maintain strict and frequent adherence to glucose supplementation through day and night, by way of a slow glucose release formula. One missed or delayed dose can result in emergency hypoglycemia. Among many complications, enlarged liver, accumulation of uric acid, lactate, and lipids are common in GSD-Ia patients.

FIG. 5 shows a schematic illustrating that base editors as described herein generate permanent, predicted nucleotide substitutions in an editing window. The R83C mutation introduces a single G>A conversion in the *g6pc* gene. Adenine base editors (ABEs) enable the programmable conversion of A to G in genomic DNA and thus may be used to correct this mutation. **FIG. 5** depicts the utility of ABEs and base editing as described herein. ABE

binds to target DNA that is complementary to the guide-RNA and exposes a stretch of single-stranded DNA. The deaminase converts the target adenine into inosine, and the Cas enzyme nicks the opposite strand, which is then repaired, completing the base pair conversion. The direct repair of a point mutation has the potential for restoration of gene function.

FIGs. 6A and 6B provide a depiction of the target nucleotide site, and bystander and PAM nucleotides and a bar graph showing that ABEs used in immortalized HEK293 cells yield a significant rate of precise correction of R83C. Base-editors for A>G conversion in the g6pc gene were optimized for correction of R83C. Shown in **FIG. 6A** is the target DNA sequence (CCACCAGTATGGACACTGTCCAAAGAGAAT (SEQ ID NO: 402)) and underlying amino acid translation (WWYPCQGFLI; SEQ ID NO: 403) for the GSD-Ia R83C mutation. The target edit is shown by double-underlining, at position 12. The editing window also includes a possible bystander, shown by single-underlining at position 6, and an edit that may result in a synonymous conversion is shown at position 10. For screening, a HEK293 cell line was generated to express the *g6pc* transgene harboring the R83C mutation and was transfected with base-editor mRNA and gRNA. Allele frequencies were assessed by high-throughput targeted amplicon Next-Generation Sequencing (NGS). Variants 1-5 represent a combination of gRNA and base-editor RNA, engineered for optimized target correction. Variant 5 yielded approximately 60% targeted base-editing efficiency for R83C correction with limited bystander editing (**FIG. 6B**).

FIG. 7 presents a photographic image and bar graphs demonstrating that 3-week-old homozygous huR83C (Hom huR83C) mice exhibited expected growth impairment and metabolic defects characteristic of GSD-1a. For the experients, a GSD-Ia mouse that expresses the human G6PC-R83C transgene in place of mouse G6PC was generated to validate base-editing *in vivo*. The results shown confirmed that mice homozygous for huR83C exhibited postnatal lethality -- they were either stillborn or died within 24 hours. On glucose supplementation therapy, the animals survived to at least 3 weeks of age and revealed characteristic pathological signatures of GSD-Ia, with reduced body weight, enlarged livers, significant G6Pase inhibition, and abnormal serum metabolites as compared to littermate controls, a phenotype that is consistent with clinical and published reports.

FIGs. 8A and 8B show dot plots of *in vivo* correction achieved by the base editors (ABEs) described herein. **FIG. 8A** illustrates efficient lipid nanoparticle (LNP)-mediated base editing (huG6PC-R83C correction) in livers of adult and newborn heterozygous huR83C mice. To validate base-editing efficiency for R83C correction in vivo, LNP-mediated

delivery was first optimized in less fragile transgenic mice heterozygous for huR83C. The schematic in FIG. 2A depicts *in vivo* workflow for these experiments, with lipid nanoparticle (LNP), or LNP co-formulations of base-editor mRNA and gRNA dosed via IV injection. Given neonatal lethality of the homozygous mice, LNP-dosing was employed via the temporal vein of heterozygous huR83C mice shortly post birth, and activity was compared to that seen in adult heterozygous huR83C mice that had received LNP administered via the tail vein. NGS analysis of whole liver extracts revealed approximately 40% base-editing efficiency in adults and up to ~60% efficiency in newborns, with a broader range in efficiencies. Bystander editing remained low in adults and newborns. **FIG. 8B** shows that LNP-mediated R83C correction in livers is associated with survival of newborn homozygous huR83C mice and littermate heterozygous huR83C mice. Briefly, newborn mice homozygous for huR83C were treated with LNP containing guide RNA and mRNA encoding ABE. The treated mice grew normally to 3 weeks of age, without hypoglycemia-induced seizures, in the absence of glucose therapy. The treated homozygous huR83C mice displayed editing efficiencies up to ~60% in total liver extracts (i.e., ~60% R83C correction), consistent with littermate controls that were heterozygous for huR83C.

FIGs. 9A and 9B show bar graphs and immunohistochemical staining images demonstrating the base editing as described herein in mice homozygous for huG6PC-R83C restores near-normal metabolic function to reverse GSD-Ia pathology. At 3 weeks, it was validated that the treated homozygous huR83C mice displayed proper metabolic function, with restoration of near-normal serum metabolite markers, including glucose, triglycerides, cholesterol, lactate, and uric acid, as shown by the darkest bars in the graph in **FIG. 9A**. Moreover, biochemical assays of G6PC activity (as assessed biochemically and via lead-phosphate staining) in LNP-treated homozygous huR83C mice were consistent with that of litter-mate controls. Hepatomegaly, another clinical presentation of GSD-Ia, is caused primarily by excess glycogen and lipid deposition. Immuno-histochemical analysis revealed normal hepatocyte size and lipid deposition in LNP-treated mice (**FIG. 9B**). The results demonstrate the potential of base-editing to correct the R83C mutation and the metabolic defects associated with GSD-Ia.

FIG. 10 shows a bar graph demonstrating that a single LNP dose administration in homozygous huG6PC-R83C mice maintained euglycemia during a 24-hour fasting challenge via base-editing as described herein.

FIG. 11 shows a bar graph demonstrating the results of experiments in which representative heavy mods gRNAs (heavy mod series 1 gRNAs for R83C (saCas9), Example 5) were used in experiments in which adult transgenic mice heterozygous for huG6PC-R83C were dosed at a sub-saturating dose of 1mpk of 1:1 ratio of gRNA:editor mRNA.

FIG. 12 presents a graph illustrating Kaplan-Meier survival curves generated to estimate the survival of newborn transgenic mice homozygous for huG6PC-R83C, either post base-editing via ABE mRNA (ABE-treated) or untreated (Untreated). The top line plot on the graph represents the survival of animals following base-editing via ABE mRNA (ABE-treated), (100% survival over time (3 wks)), as described herein (Example 2). The leftmost line plot on the graph represents the survival of untreated animals over time (poor to no survival at less than 1 week).

DETAILED DESCRIPTION OF THE EMBODIMENTS

Provided and featured herein are compositions comprising novel adenosine base editors that have increased efficiency and methods of using base editors comprising adenosine deaminase variants for altering mutations associated with Glycogen Storage Disease Type 1a (GSD1a).

The embodiments described herein are based, at least in part, on the discovery that a base editor featuring adenosine deaminase variants (*e.g.*, adenosine deaminase variants that comprise a combination of alterations in a TadA*7.10 amino acid sequence, where the combination of alterations is V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, or a corresponding combination of alterations in another adenosine deaminase) precisely corrects single nucleotide polymorphisms in the endogenous glucose-6-phosphatase (G6PC) gene (*e.g.* R83C, Q347X).

The GSD1a mutations, R83C and Q347X, are cytidine to thymidine (C→T) transition mutations, resulting in a C•G to T•A base pair substitution. These substitutions may be reverted back to a wild-type, non-pathogenic genomic sequence with an adenosine base editor (ABE) which catalyzes A•T to G•C substitutions. By extension, GSD1a-causing mutations are potential targets for reversion to wild-type sequence using ABEs without the risks of inducing G6PC gene overexpression, as may occur using gene therapy. Accordingly, A•T to G•C DNA base editing precisely corrects one or more of the most prevalent GSD1a-causing mutations in the G6PC gene.

NUCLEOBASE EDITORS

Useful in the methods and compositions described herein are nucleobase editors that edit, modify or alter a target nucleotide sequence of a polynucleotide. Nucleobase editors described herein typically include a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, adenosine deaminase or cytidine deaminase). A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target polynucleotide sequence and thereby localize the base editor to the target nucleic acid sequence desired to be edited.

In certain embodiments, the nucleobase editors provided herein comprise one or more features that improve base editing activity. For example, any of the nucleobase editors provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the nucleobase editors provided herein may have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9). Without wishing to be bound by any particular theory, the presence of the catalytic residue (*e.g.*, H840) maintains the activity of the Cas9 to cleave the non-edited (*e.g.*, non-deaminated) strand opposite the targeted nucleobase. Mutation of the catalytic residue (*e.g.*, D10 to A10) prevents cleavage of the edited (*e.g.*, deaminated) strand containing the targeted residue (*e.g.*, A or C). Such Cas9 variants can generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a nucleobase change on the non-edited strand.

Polynucleotide Programmable Nucleotide Binding Domain

Polynucleotide programmable nucleotide binding domains bind polynucleotides (*e.g.*, RNA, DNA). A polynucleotide programmable nucleotide binding domain of a base editor can itself comprise one or more domains (*e.g.*, one or more nuclease domains). In some embodiments, the nuclease domain of a polynucleotide programmable nucleotide binding domain can comprise an endonuclease or an exonuclease. An endonuclease can cleave a single strand of a double-stranded nucleic acid or both strands of a double-stranded nucleic acid molecule. In some embodiments, a nuclease domain of a polynucleotide programmable nucleotide binding domain can cut zero, one, or two strands of a target polynucleotide.

Non-limiting examples of a polynucleotide programmable nucleotide binding domain which can be incorporated into a base editor include a CRISPR protein-derived domain, a

restriction nuclease, a meganuclease, TAL nuclease (TALEN), and a zinc finger nuclease (ZFN). In some embodiments, a base editor comprises a polynucleotide programmable nucleotide binding domain comprising a natural or modified protein or portion thereof which via a bound guide nucleic acid is capable of binding to a nucleic acid sequence during CRISPR (*i.e.*, Clustered Regularly Interspaced Short Palindromic Repeats)-mediated modification of a nucleic acid. Such a protein is referred to herein as a "CRISPR protein." Accordingly, disclosed herein is a base editor comprising a polynucleotide programmable nucleotide binding domain comprising all or a portion of a CRISPR protein (*i.e.* a base editor comprising as a domain all or a portion of a CRISPR protein, also referred to as a "CRISPR protein-derived domain" of the base editor). A CRISPR protein-derived domain incorporated into a base editor can be modified compared to a wild-type or natural version of the CRISPR protein. For example, as described below a CRISPR protein-derived domain can comprise one or more mutations, insertions, deletions, rearrangements and/or recombinations relative to a wild-type or natural version of the CRISPR protein.

Cas proteins that can be used herein include class 1 and class 2. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 or Csx12), Cas10, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Cas12a/Cpf1, Cas12b/C2c1 (*e.g.*, SEQ ID NO: 232), Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, and Cas12j/CasΦ, CARF, DinG, homologues thereof, or modified versions thereof. A CRISPR enzyme can direct cleavage of one or both strands at a target sequence, such as within a target sequence and/or within a complement of a target sequence. For example, a CRISPR enzyme can direct cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

A vector that encodes a CRISPR enzyme that is mutated to with respect, to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence can be used. A Cas protein (*e.g.*, Cas9, Cas12) or a Cas domain (*e.g.*, Cas9, Cas12) can refer to a polypeptide or domain with at least or at least about 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity and/or sequence

homology to a wild-type exemplary Cas polypeptide or Cas domain. Cas (*e.g.*, Cas9, Cas12) can refer to the wild-type or a modified form of the Cas protein that can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof.

In some embodiments, a CRISPR protein-derived domain of a base editor can include all or a portion of Cas9 from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1), *Streptococcus pyogenes*, or *Staphylococcus aureus*.

Cas9 nuclease sequences and structures are well known to those of skill in the art (See, *e.g.*, "Complete genome sequence of an Ml strain of Streptococcus pyogenes." Ferretti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., *et al.*, Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., *et al.*, *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, S. pyogenes and S. thermophilus. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference.

In some embodiments, the gRNA scaffold sequence is as follows:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG CACCGAGUCGGUGCUUUU (SEQ ID NO: 317).

In an embodiment, the RNA scaffold comprises a stem loop. In an embodiment, the RNA scaffold comprises the nucleic acid sequence:

GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAACGAAACUUACACAGUUACUUAAAUCUU

GCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGU
G (SEQ ID NO: 389).

In an embodiment, the RNA scaffold comprises a canonical stem loop. In an embodiment, the RNA scaffold comprises the nucleic acid sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCU*mU*mU*mU (SEQ ID NO: 324)

where m=2'-O-methyl modification and *=3' phosphorothioate internucleotide linkages (*i.e.*, at the first 3' terminal RNA residues as shown here).

In an embodiment, the RNA scaffold comprises the nucleic acid sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGCUUUU (SEQ ID NO: 390).

In an embodiment, an *S. pyogenes* sgRNA scaffold polynucleotide sequence is as follows:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGC (SEQ ID NO: 319).

In an embodiment, an *S. aureus* sgRNA scaffold polynucleotide sequence is as follows:

GUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUA
UCUCGUCAACUUGUUGGCGAGA (SEQ ID NO: 320).

In an embodiment, the RNA scaffold comprises a non-canonical sequence. In an embodiment, the RNA scaffold comprises the nucleic acid sequence:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
GACCGAGUCGGUGCU*mU*mU*mU (SEQ ID NO: 323)

where m=2'-O-methyl modification and *=3' phosphorothioate internucleotide linkages (*i.e.*, at the first 3' terminal RNA residues as shown here).

In some embodiments, wild-type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1). An exemplary *Streptococcus pyogenes* Cas9 (spCas9) nucleic acid sequence is provided below:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGAT
CACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACA
GTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACT
CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACA
GGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGT

CTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT

GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTC

TACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTG

GTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATC

CAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTAGA

TGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTC

AGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTG

ACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGA

TACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGT

TTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGT

GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGA

CTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTT

TTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT

TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACT

AAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAA

TTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAA

GACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATT

GGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCAT

GGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACA

AACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTA

TTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAG

CATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAA

GTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGA

AATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAA

TTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTT

TTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAGACTTAAAACATATGCTCA

CCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTT

TGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTT

TTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC

ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGA

TTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTT

GATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGA

AAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAG

GTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAA

AATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATT
AGATATTAATCGTTTAAGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAG
ACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAAC
GTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAA
GTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC
TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTG
GCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA
GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCT
ATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTT
GGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAA
AGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAA
AATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGA
GAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA
AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGA
AAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAC
AAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAAC
GGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAAT
CCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATT
GACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAA
ATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTAC
AAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCAT
TATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCA
TAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAG
CAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGT
GAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTT
TAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATG
CCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTA
GGAGGTGACTGA (SEQ ID NO: 198).

An exemplary *Streptococcus pyogenes* Cas9 (spCas9) amino acid sequence is provided below:

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI

QLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNS
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIV
DELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD
KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD (SEQ ID NO: 199)

(single underline: HNH domain; double underline: RuvC domain)

In some embodiments, wild-type Cas9 corresponds to, or comprises the following nucleotide sequences:

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCAT
AACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATT
CGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACT
CGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACA
AGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGT
CCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTC
AACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTG
GGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATC
CAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGA

TGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCAC

AATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTG

ACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGA

CACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTAT

TTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACT

GAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGA

CTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCT

TTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTC

TACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACT

CAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAA

TCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAA

GACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCT

GGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCAT

GGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACC

AACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTA

TTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCG

CCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAA

GTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGA

GATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGA

TAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTG

TTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCA

CCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGAT

TGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTT

CTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC

CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATA

TTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTG

GATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACG

CGAAAATCAAACGACTCAGAAGGGGCAAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAG

AGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTG

CAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGA

ACTGGACATAAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGA

AGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGAC

AATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGC

GAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG

AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCAT
GTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCG
GGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAAT
TCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTC
GTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTA
CAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAG
CCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAAC
GGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGA
TAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAA
AGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGT
GATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCC
TACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGA
AGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCC
ATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACC
AAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGC
TTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCC
CATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCA
GCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCC
TAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATA
CGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGC
ATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAG
ACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAG
CTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGA
CGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

(SEQ ID NO: 200)

In some embodiments, wild-type Cas9 corresponds to, or comprises the following amino acid sequence:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK

DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD (SEQ ID NO: 201).

(single underline: HNH domain; double underline: RuvC domain).

In some embodiments, wild-type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_002737.2.In some embodiments, wild-type Cas9 corresponds to Cas9 from *Streptococcuspyogenes* (Uniprot Reference Sequence: Q99ZW2).

The amino acid sequence of an exemplary catalytically inactive Cas9 (dCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF

LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP
IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS
HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI
REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD (SEQ ID NO: 203) (see, *e.g.*, Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference).

In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an "nCas9" protein (for "nickase" Cas9). A nuclease-inactivated Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9) or catalytically inactive Cas9. Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (*See, e.g.*, Jinek *et al., Science*. 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell*. 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al., Science*. 337:816-821(2012); Qi *et al., Cell*. 28;152(5):1173-83 (2013)).

Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and

D10A/D839A/H840A/N863A mutant domains (*See*, *e.g.*, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference).

In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase, referred to as an "nCas9" protein (for "nickase" Cas9). The Cas9 nickase may be a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule (*e.g.*, a duplexed DNA molecule). In some embodiments the Cas9 nickase cleaves the target strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is base paired to (complementary to) a gRNA (*e.g.*, an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position 840.

The amino acid sequence of an exemplary catalytically Cas9 nickase (nCas9) is as follows:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEAT
RLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFI
QLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLK
DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT
NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK
VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV
LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDF
LKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNP

IDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLAS HYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPI REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGD (SEQ ID NO: 233).

In some embodiments, Cas9 is a variant Cas9 protein. A variant Cas9 polypeptide has an amino acid sequence that is different by one amino acid (*e.g.*, has a deletion, insertion, substitution, fusion) when compared to the amino acid sequence of a wild-type Cas9 protein. In some instances, the variant Cas9 polypeptide has an amino acid change (*e.g.*, deletion, insertion, or substitution) that reduces the nuclease activity of the Cas9 polypeptide. For example, in some instances, the variant Cas9 polypeptide has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9 protein. In some embodiments, the variant Cas9 protein has no substantial nuclease activity.

In some embodiments the Cas9 nickase cleaves the non-target, non-base-edited strand of a duplexed nucleic acid molecule, meaning that the Cas9 nickase cleaves the strand that is not base paired to a gRNA (e.g., an sgRNA) that is bound to the Cas9. In some embodiments, a Cas9 nickase comprises an H840A mutation and has an aspartic acid residue at position 10, or a corresponding mutation. In some embodiments the Cas9 nickase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the Cas9 nickases provided herein. Additional suitable Cas9 nickases will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure.

In some embodiments, Cas9 is a modified Cas9. A given gRNA targeting sequence can have additional sites throughout the genome where partial homology exists. These sites are called off-targets and need to be considered when designing a gRNA. In addition to optimizing gRNA design, CRISPR specificity can also be increased through modifications to Cas9. Cas9 generates double-strand breaks (DSBs) through the combined activity of two nuclease domains, RuvC and HNH. Cas9 nickase, a D10A mutant of SpCas9, retains one nuclease domain and generates a DNA nick rather than a DSB. The nickase system can also be combined with HDR-mediated gene editing for specific gene edits.

Catalytically Dead Nucleases

Also provided herein are base editors comprising a polynucleotide programmable nucleotide binding domain which is catalytically dead (*i.e.*, incapable of cleaving a target polynucleotide sequence). Herein the terms "catalytically dead" and "nuclease dead" are used interchangeably to refer to a polynucleotide programmable nucleotide binding domain which has one or more mutations and/or deletions resulting in its inability to cleave a strand of a nucleic acid. In some embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain base editor can lack nuclease activity as a result of specific point mutations in one or more nuclease domains. For example, in the case of a base editor comprising a Cas9 domain, the Cas9 can comprise both a D10A mutation and an H840A mutation. Such mutations inactivate both nuclease domains, thereby resulting in the loss of nuclease activity. In other embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain can comprise one or more deletions of all or a portion of a catalytic domain (*e.g.*, RuvC1 and/or HNH domains). In further embodiments, a catalytically dead polynucleotide programmable nucleotide binding domain comprises a point mutation (*e.g.*, D10A or H840A) as well as a deletion of all or a portion of a nuclease domain. dCas9 domains are known in the art and described, for example, in Qi *et al.*, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." *Cell*. 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference.

Additional suitable nuclease-inactive dCas9 domains will be apparent to those of skill in the art based on this disclosure and knowledge in the field, and are within the scope of this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A/H840A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, *e.g.*, Prashant *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*. 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference).

In some embodiments, dCas9 corresponds to, or comprises in part or in whole, a Cas9 amino acid sequence having one or more mutations that inactivate the Cas9 nuclease activity. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10X mutation and a H840X mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid change. In some embodiments, the nuclease-inactive dCas9 domain comprises a D10A mutation and a

H840A mutation of the amino acid sequence set forth herein, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, a nuclease-inactive Cas9 domain comprises the amino acid sequence set forth in Cloning vector pPlatTET-gRNA2 (Accession No. BAV54124).

In some embodiments, a variant Cas9 protein can cleave the complementary strand of a guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the RuvC domain. As a non-limiting example, in some embodiments, a variant Cas9 protein has a D10A (aspartate to alanine at amino acid position 10) and can therefore cleave the complementary strand of a double stranded guide target sequence but has reduced ability to cleave the non-complementary strand of a double stranded guide target sequence (thus resulting in a single strand break (SSB) instead of a double strand break (DSB) when the variant Cas9 protein cleaves a double stranded target nucleic acid) (see, for example, Jinek *et al.*, *Science*. 2012 Aug. 17; 337(6096):816-21).

In some embodiments, a variant Cas9 protein can cleave the non-complementary strand of a double stranded guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus resulting in a SSB instead of a DSB when the variant Cas9 protein cleaves a double stranded guide target sequence). Such a Cas9 protein has a reduced ability to cleave a guide target sequence (*e.g.*, a single stranded guide target sequence) but retains the ability to bind a guide target sequence (*e.g.*, a single stranded guide target sequence).

In some embodiments, a variant Cas9 protein can cleave the non-complementary strand of a double stranded guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence. For example, the variant Cas9 protein can have a mutation (amino acid substitution) that reduces the function of the HNH domain (RuvC/HNH/RuvC domain motifs). As a non-limiting example, in some embodiments, the variant Cas9 protein has an H840A (histidine to alanine at amino acid position 840) mutation

and can therefore cleave the non-complementary strand of the guide target sequence but has reduced ability to cleave the complementary strand of the guide target sequence (thus resulting in a SSB instead of a DSB when the variant Cas9 protein cleaves a double stranded guide target sequence). Such a Cas9 protein has a reduced ability to cleave a guide target sequence (e.g., a single stranded guide target sequence) but retains the ability to bind a guide target sequence (e.g., a single stranded guide target sequence).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors W476A and W1126A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (e.g., a single stranded target DNA) but retains the ability to bind a target DNA (e.g., a single stranded target DNA).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (e.g., a single stranded target DNA) but retains the ability to bind a target DNA (e.g., a single stranded target DNA).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors H840A, W476A, and W1126A, mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (e.g., a single stranded target DNA) but retains the ability to bind a target DNA (e.g., a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors H840A, D10A, W476A, and W1126A, mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (e.g., a single stranded target DNA) but retains the ability to bind a target DNA (e.g., a single stranded target DNA). In some embodiments, the variant Cas9 has restored catalytic His residue at position 840 in the Cas9 HNH domain (A840H).

As another non-limiting example, in some embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (e.g., a single stranded target DNA) but retains the ability to bind a target DNA (e.g., a single stranded target DNA). In some embodiments, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1127A mutations,

the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such embodiments, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some embodiments, when such a variant Cas9 protein is used in a method of binding, the method can include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (i.e., inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (i.e., substituted). Also, mutations other than alanine substitutions are suitable.

In some embodiments, the variant Cas protein can be spCas9, spCas9-VRQR, spCas9-VRER, xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER, spCas9-LRKIQK, or spCas9-LRVSQL.

In some embodiments, a modified SpCas9 including amino acid substitutions D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (SpCas9-MQKFRAER) and having specificity for the altered PAM 5'-NGC-3' was used.

In some embodiments, the Cas9 is a Cas9 variant having specificity for an altered PAM sequence. In some embodiments, the Additional Cas9 variants and PAM sequences are described in Miller, S.M., *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs, Nat. Biotechnol. (2020), the entirety of which is incorporated herein by reference. in some embodiments, a Cas9 variate have no specific PAM requirements. In some embodiments, a Cas9 variant, *e.g.* a SpCas9 variant has specificity for a NRNH PAM, wherein R is A or G and H is A, C, or T. In some embodiments, the SpCas9 variant has specificity for a PAM sequence AAA, TAA, CAA, GAA, TAT, GAT, or CAC. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1218, 1219, 1221, 1249, 1256, 1264, 1290, 1318, 1317, 1320, 1321, 1323, 1332, 1333, 1335, 1337, or 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1135, 1218, 1219, 1221, 1249, 1320, 1321, 1323, 1332, 1333, 1335, or 1337 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1256, 1264, 1290, 1318, 1317, 1320, 1323, 1333 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at

position 1114, 1131, 1135, 1150, 1156, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1286, 1293, 1320, 1321, 1332, 1335, 1339 or a corresponding position thereof. In some embodiments, the SpCas9 variant comprises an amino acid substitution at position 1114, 1127, 1135, 1180, 1207, 1219, 1234, 1286, 1301, 1332, 1335, 1337, 1338, 1349 or a corresponding position thereof. Exemplary amino acid substitutions and PAM specificity of SpCas9 variants are shown in **Tables 1A-1D**.

Table 1A

SpCas9	SpCas9 amino acid position												
	1114	1135	1218	1219	1221	1249	1320	1321	1323	1332	1333	1335	1337
	R	D	G	E	Q	P	A	P	A	D	R	R	T
AAA		N		V	H						G		
AAA		N		V	H						G		
AAA				V							G		
TAA	G	N		V							I		
TAA		N		V							I		A
TAA	G	N		V							I		A
CAA				V							K		
CAA		N		V							K		
CAA		N		V							K		
GAA				V	H		V				K		
GAA		N		V			V				K		
GAA				V	H		V				K		
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
TAT			S	V	H	S		S				L	
GAT				V							I		
GAT				V					D			Q	
GAT				V					D			Q	
CAC				V						N		Q	N
CAC		N		V								Q	N
CAC				V						N		Q	N

Table 1B

SpCas9	SpCas9 amino acid position																		
	1114	1134	1135	1137	1139	1151	1180	1188	1211	1219	1221	1256	1264	1290	1318	1317	1320	1323	1333
	R	F	D	P	V	K	D	K	K	E	Q	Q	H	V	L	N	A	A	R
GAA										V	H						V		K
GAA			N	S						V							V	D	K
GAA			N							V	H		Y				V		K
CAA			N							V	H		Y				V		K
CAA	G		N	S						V	H		Y				V		K
CAA			N					R		V	H						V		K
CAA			N				G		R	V	H		Y				V		K
CAA			N							V	H		Y				V		K
AAA			N				G			V	H	R	Y				V	D	K
CAA	G		N				G			V	H		Y				V	D	K
CAA		L	N				G			V	H		Y			T	V	D	K
TAA	G		N				G			V	H		Y	G	S		V	D	K
TAA	G		N			E	G			V	H		Y		S		V		K
TAA	G		N				G			V	H		Y		S		V	D	K
TAA	G		N				G		R	V	H						V		K
TAA			N				G		R	V	H		Y				V		K
TAA	G		N		A		G			V	H						V		K
TAA	G		N							V	H						V		K

Table 1C

SpCas9	SpCas9 amino acid position																		
	1114	1131	1135	1150	1156	1180	1191	1218	1219	1221	1227	1249	1253	1286	1293	1320	1321	1332	1335	1339
	R	Y	D	E	K	D	K	G	E	Q	A	P	E	N	A	A	P	D	R	T
SacB. TAT			N				N		V	H						V	S		L	
SacB. TAT			N					S	V	H		S					S	G	L	
AAT			N					S	V	H	V	S		K	T		S	G	L	I
TAT	G		N			G		S	V	H		S	K				S	G	L	
TAT	G		N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N		E	G		S	V	H		S					S	G	L	
TAT	G	C	N	V		G		S	V	H		S					S	G	L	
TAT		C	N			G		S	V	H		S					S	G	L	
TAT	G	C	N			G		S	V	H		S					S	G	L	

Table 1D

SpCas9	SpCas9 amino acid position													
	1114	1127	1135	1180	1207	1219	1234	1286	1301	1332	1335	1337	1338	1349
	R	D	D	D	E	E	N	N	P	D	R	T	S	H
SacB.CAC			N			V				N	Q	N		
AAC	G		N			V				N	Q	N		
AAC	G		N			V				N	Q	N		
TAC	G		N			V				N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N		G	V	D	H		N	Q	N		
TAC	G		N			V				N	Q	N		
TAC	G	G	N	E		V		H		N	Q	N		
TAC	G		N			V		H		N	Q	N		
TAC	G		N			V				N	Q	N	T	R

Nucleic acid programmable DNA binding proteins

Some aspects of the disclosure provide fusion proteins comprising domains that act as nucleic acid programmable DNA binding proteins, which may be used to guide a protein, such as a base editor, to a specific nucleic acid (*e.g.*, DNA or RNA) sequence. In particular embodiments, a fusion protein comprises a nucleic acid programmable DNA binding protein domain and a deaminase domain. Non-limiting examples of nucleic acid programmable DNA binding proteins include, Cas9 (*e.g.*, dCas9 and nCas9),

In some embodiments, one of the Cas9 domains present in the fusion protein may be replaced with a guide nucleotide sequence-programmable DNA-binding protein domain that has no requirements for a PAM sequence.

In some embodiments, the Cas9 domain is a Cas9 domain from *Staphylococcus aureus* (SaCas9). In some embodiments, the SaCas9 domain is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n). In some embodiments, the SaCas9 comprises a N579A mutation, or a corresponding mutation in any of the amino acid sequences provided herein.

In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a

nucleic acid sequence having a NNGRRT or a NNGRRT PAM sequence. The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

In some embodiments, the SaCas9 domain comprises one or more of a E781X, a N967X, and a R1014X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SaCas9 domain comprises one or more of a E781K, a N967K, and a R1014H mutation, or one or more corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SaCas9 domain comprises a E781K, a N967K, or a R1014H mutation, or corresponding mutations in any of the amino acid sequences provided herein.

Exemplary SaCas9 sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEE**N**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG

(SEQ ID NO: 218).

Residue N579 above, which is underlined and in bold, may be mutated (*e.g.*, to a A579) to yield a SaCas9 nickase.

Exemplary SaCas9n sequence

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHR
IQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVE
EDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQ
KAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKY
AYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIK
GYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNS
ELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKE
IPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKR
NRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHII
PRSVSFDNSFNNKVLVKQEE**A**SKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKT
KKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSF
LRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIE
TEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNG
LYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYS
KKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNL
DVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIE
VNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
(SEQ ID NO: 219).

Residue A579 above, which can be mutated from N579 to yield a SaCas9 nickase, is underlined and in bold.

In some embodiments, the napDNAbp is a circular permutant (e.g., SEQ ID NO: 238). In the following sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence, and double underlined sequence indicates mutations. The asterisk (*) denotes a STOP codon.

CP5 (with MSP "NGC=Pam Variant with mutations Regular Cas9 likes NGG" PID=Protein Interacting Domain and "D10A" nickase):

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPTVAYSVLVVAKVEK
GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE
FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIARKEYR

STKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGSGGM*DKKYSIGLAI
GTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI
YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA
EDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM
DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILT
FRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV
LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYF
KKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHK
PENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNT
KYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK
LESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV* (SEQ ID NO: 238).

Single effectors of microbial CRISPR-Cas systems include, without limitation, Cas9, Cpf1, Cas12b/C2c1, and Cas12c/C2c3. Typically, microbial CRISPR-Cas systems are divided into Class 1 and Class 2 systems. Class 1 systems have multisubunit effector complexes, while Class 2 systems have a single protein effector. For example, Cas9 and Cpf1 are Class 2 effectors. In addition to Cas9 and Cpf1, three distinct Class 2 CRISPR-Cas systems (Cas12b/C2c1, and Cas12c/C2c3) have been described by Shmakov *et al.*, "Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems", *Mol. Cell,* 2015 Nov. 5; 60(3): 385-397, the entire contents of which is hereby incorporated by reference. Effectors of two of the systems, Cas12b/C2c1, and Cas12c/C2c3, contain RuvC-like endonuclease domains related to Cpf1. A third system contains an effector with two predicated HEPN RNase domains. Production of mature CRISPR RNA is tracrRNA-independent, unlike production of CRISPR RNA by Cas12b/C2c1. Cas12b/C2c1 depends on both CRISPR RNA and tracrRNA for DNA cleavage.

In some embodiments, the napDNAbp is a circular permutant (*e.g.*, SEQ ID NO: 238).

The crystal structure of *Alicyclobaccillus acidoterrastris* Cas12b/C2c1 (AacC2c1) has been reported in complex with a chimeric single-molecule guide RNA (sgRNA). See *e.g.*, Liu *et al.*, "C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism", *Mol. Cell,* 2017 Jan. 19; 65(2):310-322, the entire contents of which are hereby incorporated by reference. The crystal structure has also been reported in *Alicyclobacillus acidoterrestris* C2c1 bound to target DNAs as ternary complexes. See *e.g.*, Yang *et al.*, "PAM-dependent Target DNA Recognition and Cleavage by C2C1 CRISPR-Cas endonuclease", *Cell,* 2016 Dec. 15; 167(7):1814-1828, the entire contents of which are hereby incorporated by reference. Catalytically competent conformations of AacC2c1, both with target and non-target DNA strands, have been captured independently positioned within a single RuvC catalytic pocket, with Cas12b/C2c1-mediated cleavage resulting in a staggered seven-nucleotide break of target DNA. Structural comparisons between Cas12b/C2c1 ternary complexes and previously identified Cas9 and Cpf1 counterparts demonstrate the diversity of mechanisms used by CRISPR-Cas9 systems.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12b/C2c1, or a Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a Cas12b/C2c1 protein. In some embodiments, the napDNAbp is a Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12b/C2c1 or Cas12c/C2c3 protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any one of the napDNAbp sequences provided herein. It should be appreciated that Cas12b/C2c1 or Cas12c/C2c3 from other bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12c. In some embodiments, the Cas12c protein is a Cas12c1 (SEQ ID NO: 239) or a variant of Cas12c1. In some embodiments, the Cas12 protein is a Cas12c2 (SEQ ID NO: 240) or a variant of Cas12c2. In some embodiments, the Cas12 protein is a Cas12c protein from *Oleiphilus* sp. HI0009 (*i.e.*, OspCas12c; SEQ ID NO: 241) or a variant of OspCas12c. These Cas12c molecules have

been described in Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91; the entire contents of which is hereby incorporated by reference. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12c1, Cas12c2, or OspCas12c protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any Cas12c1, Cas12c2, or OspCas12c protein described herein. It should be appreciated that Cas12c1, Cas12c2, or OspCas12c from other bacterial species may also be used in accordance with the present disclosure.

In some embodiments, a napDNAbp refers to Cas12g, Cas12h, or Cas12i, which have been described in, for example, Yan *et al.*, "Functionally Diverse Type V CRISPR-Cas Systems," Science, 2019 Jan. 4; 363: 88-91; the entire contents of each is hereby incorporated by reference. Exemplary Cas12g, Cas12h, and Cas12i polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 242-245. By aggregating more than 10 terabytes of sequence data, new classifications of Type V Cas proteins were identified that showed weak similarity to previously characterized Class V protein, including Cas12g, Cas12h, and Cas12i. In some embodiments, the Cas12 protein is a Cas12g or a variant of Cas12g. In some embodiments, the Cas12 protein is a Cas12h or a variant of Cas12h. In some embodiments, the Cas12 protein is a Cas12i or a variant of Cas12i. It should be appreciated that other RNA-guided DNA binding proteins may be used as a napDNAbp, and are within the scope of this disclosure. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12g, Cas12h, or Cas12i protein. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to any Cas12g, Cas12h, or Cas12i protein described herein. It should be appreciated that Cas12g, Cas12h, or Cas12i from other

bacterial species may also be used in accordance with the present disclosure. In some embodiments, the Cas12i is a Cas12i1 or a Cas12i2.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12j/CasΦ protein. Cas12j/CasΦ is described in Pausch *et al.*, "CRISPR-CasΦ from huge phages is a hypercompact genome editor," *Science*, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337, which is incorporated herein by reference in its entirety. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a nuclease inactive ("dead") Cas12j/CasΦ protein. It should be appreciated that Cas12j/CasΦ from other species may also be used in accordance with the present disclosure.

In some embodiments, the nucleic acid programmable DNA binding protein (napDNAbp) of any of the fusion proteins provided herein may be a Cas12j/CasΦ protein. Cas12j/CasΦ is described in Pausch et al., "CRISPR-CasΦ from huge phages is a hypercompact genome editor," Science, 17 July 2020, Vol. 369, Issue 6501, pp. 333-337, which is incorporated herein by reference in its entirety. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at ease 99.5% identical to a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a naturally-occurring Cas12j/CasΦ protein. In some embodiments, the napDNAbp is a nuclease inactive ("dead") Cas12j/CasΦ protein. It should be appreciated that Cas12j/CasΦ from other species may also be used in accordance with the present disclosure.

Guide Polynucleotides

A polynucleotide programmable nucleotide binding domain, when in conjunction with a bound guide polynucleotide (*e.g.*, gRNA), can specifically bind to a target polynucleotide sequence (*i.e.*, via complementary base pairing between bases of the bound guide nucleic acid and bases of the target polynucleotide sequence) and thereby localize the base editor to the target nucleic acid sequence desired to be edited. In some embodiments,

the target polynucleotide sequence comprises single-stranded DNA or double-stranded DNA. In some embodiments, the target polynucleotide sequence comprises RNA. In some embodiments, the target polynucleotide sequence comprises a DNA-RNA hybrid.

In an embodiment, the guide polynucleotide is a guide RNA. Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA," or simply "gRNA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. *See, e.g.*, Jinek M. *et al.*, Science 337:816-821(2012), the entire contents of which is hereby incorporated by reference.

In some embodiments, the guide polynucleotide is at least one single guide RNA ("sgRNA" or "gRNA"). In some embodiments, the guide polynucleotide is at least one tracrRNA. In some embodiments, the guide polynucleotide does not require PAM sequence to guide the polynucleotide-programmable DNA-binding domain (*e.g.*, Cas9) to the target nucleotide sequence.

A guide polynucleotide can be DNA. A guide polynucleotide can be RNA. As will be appreciated by one having skill in the art, in a guide polynucleotide sequence uracil (U) replaces thymine (T) in the sequence. In some embodiments, the guide polynucleotide comprises natural nucleotides (*e.g.*, adenosine). In some embodiments, the guide polynucleotide comprises non-natural (or unnatural) nucleotides (*e.g.*, peptide nucleic acid or nucleotide analogs). In some embodiments, the targeting region of a guide nucleic acid sequence can be at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A targeting region of a guide nucleic acid can be between 10-30 nucleotides in length, or between 15-25 nucleotides in length, or between 15-20 nucleotides in length. In some embodiments, a guide polynucleotide may be truncated by 1, 2, 3, 4, etc. nucleotides, particularly at the 5′ end. By way of nonlimiting example, a guide polynucleotide of 20 nucleotides in length may be truncated by 1, 2, 3, 4, etc. nucleotides, particularly at the 5′ end.

In some embodiments, a guide polynucleotide comprises two or more individual polynucleotides, which can interact with one another via for example complementary base pairing (*e.g.*, a dual guide polynucleotide). For example, a guide polynucleotide can comprise a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). For

example, a guide polynucleotide can comprise one or more trans-activating CRISPR RNA (tracrRNA).

In type II CRISPR systems, targeting of a nucleic acid by a CRISPR protein (*e.g.*, Cas9) typically requires complementary base pairing between a first RNA molecule (crRNA) comprising a sequence that recognizes the target sequence and a second RNA molecule (trRNA) comprising repeat sequences which forms a scaffold region that stabilizes the guide RNA-CRISPR protein complex. Such dual guide RNA systems can be employed as a guide polynucleotide to direct the base editors disclosed herein to a target polynucleotide sequence.

In some embodiments, the base editor provided herein utilizes a single guide polynucleotide (*e.g.*, sgRNA). In some embodiments, the base editor provided herein utilizes a dual guide polynucleotide (*e.g.*, dual gRNAs). In some embodiments, the base editor provided herein utilizes one or more guide polynucleotide (*e.g.*, multiple gRNA). In some embodiments, a single guide polynucleotide is utilized for different base editors described herein. For example, a single guide polynucleotide can be utilized for an adenosine base editor.

In some embodiments, the methods described herein can utilize an engineered Cas protein. A guide RNA (gRNA) is a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified. Exemplary gRNA scaffold sequences are provided in the sequence listing as SEQ ID NOs: 317-327. Thus, a skilled artisan can change the genomic target of the Cas protein specificity is partially determined by how specific the gRNA targeting sequence is for the genomic target compared to the rest of the genome.

In other embodiments, a guide polynucleotide can comprise both the polynucleotide targeting portion of the nucleic acid and the scaffold portion of the nucleic acid in a single molecule (*i.e.*, a single-molecule guide nucleic acid). For example, a single-molecule guide polynucleotide can be a single guide RNA (sgRNA or gRNA). Herein the term guide polynucleotide sequence contemplates any single, dual or multi-molecule nucleic acid capable of interacting with and directing a base editor to a target polynucleotide sequence.

Typically, a guide polynucleotide (*e.g.*, crRNA/trRNA complex or a gRNA) comprises a "polynucleotide-targeting segment" that includes a sequence capable of recognizing and binding to a target polynucleotide sequence, and a "protein-binding segment" that stabilizes the guide polynucleotide within a polynucleotide programmable nucleotide binding domain component of a base editor. In some embodiments, the

polynucleotide targeting segment of the guide polynucleotide recognizes and binds to a DNA polynucleotide, thereby facilitating the editing of a base in DNA. In other embodiments, the polynucleotide targeting segment of the guide polynucleotide recognizes and binds to an RNA polynucleotide, thereby facilitating the editing of a base in RNA. Herein a "segment" refers to a section or region of a molecule, *e.g.*, a contiguous stretch of nucleotides in the guide polynucleotide. A segment can also refer to a region/section of a complex such that a segment can comprise regions of more than one molecule. For example, where a guide polynucleotide comprises multiple nucleic acid molecules, the protein-binding segment of can include all or a portion of multiple separate molecules that are for instance hybridized along a region of complementarity. In some embodiments, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of "segment," unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and can include regions of RNA molecules that are of any total length and can include regions with complementarity to other molecules.

A guide RNA or a guide polynucleotide can comprise two or more RNAs, *e.g.*, CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA). A guide RNA or a guide polynucleotide can sometimes comprise a single-chain RNA, or single guide RNA (sgRNA) formed by fusion of a portion (*e.g.*, a functional portion) of crRNA and tracrRNA. A guide RNA or a guide polynucleotide can also be a dual RNA comprising a crRNA and a tracrRNA. Furthermore, a crRNA can hybridize with a target DNA.

As discussed above, a guide RNA or a guide polynucleotide can be an expression product. For example, a DNA that encodes a guide RNA can be a vector comprising a sequence coding for the guide RNA. A guide RNA or a guide polynucleotide can be transferred into a cell by transfecting the cell with an isolated guide RNA or plasmid DNA comprising a sequence coding for the guide RNA and a promoter. A guide RNA or a guide polynucleotide can also be transferred into a cell in other way, such as using virus-mediated gene delivery.

A guide RNA or a guide polynucleotide can be isolated. For example, a guide RNA can be transfected in the form of an isolated RNA into a cell or organism. A guide RNA can

be prepared by *in vitro* transcription using any *in vitro* transcription system known in the art. A guide RNA can be transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a guide RNA.

A guide RNA or a guide polynucleotide can comprise three regions: a first region at the 5' end that can be complementary to a target site in a chromosomal sequence, a second internal region that can form a stem loop structure, and a third 3' region that can be single-stranded. A first region of each guide RNA can also be different such that each guide RNA guides a fusion protein to a specific target site. Further, second and third regions of each guide RNA can be identical in all guide RNAs.

A first region of a guide RNA or a guide polynucleotide can be complementary to sequence at a target site in a chromosomal sequence such that the first region of the guide RNA can base pair with the target site. In some embodiments, a first region of a guide RNA can comprise from or from about 10 nucleotides to 25 nucleotides (*i.e.*, from 10 nucleotides to nucleotides; or from about 10 nucleotides to about 25 nucleotides; or from 10 nucleotides to about 25 nucleotides; or from about 10 nucleotides to 25 nucleotides) or more. For example, a region of base pairing between a first region of a guide RNA and a target site in a chromosomal sequence can be or can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more nucleotides in length. In some embodiments, a first region of a guide RNA can be or can be about 19, 20, or 21 nucleotides in length.

A guide RNA or a guide polynucleotide can also comprise a second region that forms a secondary structure. For example, a secondary structure formed by a guide RNA can comprise a stem (or hairpin) and a loop. A length of a loop and a stem can vary. For example, a loop can range from or from about 3 to 10 nucleotides in length, and a stem can range from or from about 6 to 20 base pairs in length. A stem can comprise one or more bulges of 1 to 10 or about 10 nucleotides. The overall length of a second region can range from or from about 16 to 60 nucleotides in length. For example, a loop can be or can be about 4 nucleotides in length and a stem can be or can be about 12 base pairs.

A guide RNA or a guide polynucleotide can also comprise a third region at the 3' end that can be essentially single-stranded. For example, a third region is sometimes not complementarity to any chromosomal sequence in a cell of interest and is sometimes not complementarity to the rest of a guide RNA. Further, the length of a third region can vary. A third region can be more than or more than about 4 nucleotides in length. For example, the length of a third region can range from or from about 5 to 60 nucleotides in length.

A guide RNA or a guide polynucleotide can target any exon or intron of a gene target. In some embodiments, a guide can target exon 1 or 2 of a gene; in other embodiments, a guide can target exon 3 or 4 of a gene. A composition can comprise multiple guide RNAs that all target the same exon or in some embodiments, multiple guide RNAs that can target different exons. An exon and an intron of a gene can be targeted.

A guide RNA or a guide polynucleotide can target a nucleic acid sequence of or of about 20 nucleotides. A target nucleic acid can be less than or less than about 20 nucleotides. A target nucleic acid can be at least or at least about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, or anywhere between 1-100 nucleotides in length. A target nucleic acid can be at most or at most about 5, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 40, 50, or anywhere between 1-100 nucleotides in length. A target nucleic acid sequence can be or can be about 20 bases immediately 5' of the first nucleotide of the PAM. A guide RNA can target a nucleic acid sequence. A target nucleic acid can be at least or at least about 1-10, 1-20, 1-30, 1-40, 1-50, 1-60, 1-70, 1-80, 1-90, or 1-100 nucleotides.

A guide polynucleotide, for example, a guide RNA, can refer to a nucleic acid that can hybridize to another nucleic acid, for example, the target nucleic acid or protospacer in a genome of a cell. A guide polynucleotide can be RNA. A guide polynucleotide can be DNA. The guide polynucleotide can be programmed or designed to bind to a sequence of nucleic acid site-specifically. A guide polynucleotide can comprise a polynucleotide chain and can be called a single guide polynucleotide. A guide polynucleotide can comprise two polynucleotide chains and can be called a double guide polynucleotide. A guide RNA can be introduced into a cell or embryo as an RNA molecule. For example, an RNA molecule can be transcribed *in vitro* and/or can be chemically synthesized. An RNA can be transcribed from a synthetic DNA molecule, *e.g.*, a gBlocks® gene fragment. A guide RNA can then be introduced into a cell or embryo as an RNA molecule. A guide RNA can also be introduced into a cell or embryo in the form of a non-RNA nucleic acid molecule, *e.g.*, a DNA molecule. For example, a DNA encoding a guide RNA can be operably linked to promoter control sequence for expression of the guide RNA in a cell or embryo of interest. An RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Plasmid vectors that can be used to express guide RNA include, but are not limited to, px330 vectors and px333 vectors. In some embodiments, a plasmid vector (*e.g.*, px333 vector) can comprise at least two guide RNA-encoding DNA sequences.

Methods for selecting, designing, and validating guide polynucleotides, *e.g.*, guide RNAs and targeting sequences are described herein and known to those skilled in the art. For example, to minimize the impact of potential substrate promiscuity of a deaminase domain in the nucleobase editor system (*e.g.*, an AID domain), the number of residues that could unintentionally be targeted for deamination (*e.g.*, off-target C residues that could potentially reside on ssDNA within the target nucleic acid locus) may be minimized. In addition, software tools can be used to optimize the gRNAs corresponding to a target nucleic acid sequence, *e.g.*, to minimize total off-target activity across the genome. For example, for each possible targeting domain choice using S. pyogenes Cas9, all off-target sequences (preceding selected PAMs, *e.g.*, NAG or NGG) may be identified across the genome that contain up to certain number (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. First regions of gRNAs complementary to a target site can be identified, and all first regions (*e.g.*, crRNAs) can be ranked according to its total predicted off-target score; the top-ranked targeting domains represent those that are likely to have the greatest on-target and the least off-target activity. Candidate targeting gRNAs can be functionally evaluated by using methods known in the art and/or as set forth herein.

As a non-limiting example, target DNA hybridizing sequences in crRNAs of a guide RNA for use with Cas9s may be identified using a DNA sequence searching algorithm. gRNA design may be carried out using custom gRNA design software based on the public tool cas-offinder as described in Bae S., Park J., & Kim J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473-1475 (2014). This software scores guides after calculating their genome-wide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24. Once the off-target sites are computationally-determined, an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In addition to identifying potential target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3 or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for a target nucleic acid sequence, *e.g.*, a target gene may be obtained and repeat elements may be screened using publicly available tools, for example, the RepeatMasker program. RepeatMasker searches input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

Following identification, first regions of guide RNAs, *e.g.*, crRNAs, may be ranked into tiers based on their distance to the target site, their orthogonality and presence of 5' nucleotides for close matches with relevant PAM sequences (for example, a 5′ G based on identification of close matches in the human genome containing a relevant PAM *e.g.*, NGG PAM for S. pyogenes, NNGRRT or NNGRRV PAM for S. aureus). As used herein, orthogonality refers to the number of sequences in the human genome that contain a minimum number of mismatches to the target sequence. A "high level of orthogonality" or "good orthogonality" may, for example, refer to 20-mer targeting domains that have no identical sequences in the human genome besides the intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality may be selected to minimize off-target DNA cleavage.

In some embodiments, a reporter system may be used for detecting base-editing activity and testing candidate guide polynucleotides. In some embodiments, a reporter system may comprise a reporter gene based assay where base editing activity leads to expression of the reporter gene. For example, a reporter system may include a reporter gene comprising a deactivated start codon, *e.g.*, a mutation on the template strand from 3'-TAC-5' to 3'-CAC-5'. Upon successful deamination of the target C, the corresponding mRNA will be transcribed as 5'-AUG-3' instead of 5'-GUG-3', enabling the translation of the reporter gene. Suitable reporter genes will be apparent to those of skill in the art. Non-limiting examples of reporter genes include gene encoding green fluorescence protein (GFP), red fluorescence protein (RFP), luciferase, secreted alkaline phosphatase (SEAP), or any other gene whose expression are detectable and apparent to those skilled in the art. The reporter system can be used to test many different gRNAs, *e.g.*, in order to determine which nucleotide residue(s) with respect to the target DNA sequence the respective deaminase will target. sgRNAs that target non-template strand nucleotide residues can also be tested in order to assess off-target effects of a specific base editing protein, *e.g.*, a Cas9 deaminase fusion protein. In some embodiments, such gRNAs can be designed such that the mutated start codon will not be base-paired with the gRNA. The guide polynucleotides can comprise standard nucleotides, modified nucleotides (*e.g.*, pseudouridine), nucleotide isomers, and/or nucleotide analogs. In some embodiments, the guide polynucleotide can comprise at least one detectable label. The detectable label can be a fluorophore (*e.g.*, FAM, TMR, Cy3, Cy5, Texas Red, Oregon Green, Alexa Fluors, Halo tags, or any other suitable fluorescent dye), a detection tag (*e.g.*, biotin, digoxigenin, and the like), quantum dots, or gold particles.

The guide polynucleotides can be synthesized chemically and/or enzymatically. For example, the guide RNA can be synthesized using standard phosphoramidite-based solid-phase synthesis methods. Alternatively, the guide RNA can be synthesized in vitro by operably linking DNA encoding the guide RNA to a promoter control sequence that is recognized by a phage RNA polymerase. Examples of suitable phage promoter sequences include T7, T3, SP6 promoter sequences, or variations thereof. In embodiments in which the guide RNA comprises two separate molecules (*e.g.*., crRNA and tracrRNA), the crRNA can be chemically synthesized and the tracrRNA can be enzymatically synthesized.

In some embodiments, a base editor system may comprise multiple guide polynucleotides, *e.g.*, gRNAs. For example, the gRNAs may target the base editor to one or more target loci (*e.g.*, at least one (1) gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 g RNA, or at least 50 gRNA). In some embodiments, multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

A DNA sequence encoding a guide RNA or a guide polynucleotide can also be part of a vector. In some embodiments, a vector comprises additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (*e.g.*, GFP or antibiotic resistance genes such as puromycin), origins of replication, and the like. A DNA molecule encoding a guide RNA or a guide polynucleotide can be circular or linear.

In some embodiments, one or more components of a base editor system may be encoded by DNA sequences. Such DNA sequences may be introduced into an expression system, *e.g.*, a cell, together or separately. For example, DNA sequences encoding a polynucleotide programmable nucleotide binding domain and a guide RNA may be introduced into a cell, each DNA sequence can be part of a separate molecule (*e.g.*, one vector containing the polynucleotide programmable nucleotide binding domain coding sequence and a second vector containing the guide RNA coding sequence) or both can be part of a same molecule (*e.g.*, one vector containing coding (and regulatory) sequence for both the polynucleotide programmable nucleotide binding domain and the guide RNA).

A guide polynucleotide can comprise one or more modifications to provide a nucleic acid with a new or enhanced feature. A guide polynucleotide can comprise a nucleic acid affinity tag. A guide polynucleotide can comprise synthetic nucleotide, synthetic nucleotide analog, nucleotide derivatives, and/or modified nucleotides.

In some embodiments, a gRNA or a guide polynucleotide can comprise modifications. A modification can be made at any location of a gRNA or a guide polynucleotide. More than one modification can be made to a single gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide can undergo quality control after a modification. In some embodiments, quality control can include PAGE, HPLC, MS, or any combination thereof.

A modification of a gRNA or a guide polynucleotide can be a substitution, insertion, deletion, chemical modification, physical modification, stabilization, purification, or any combination thereof.

A gRNA or a guide polynucleotide can also be modified by 5'adenylate, 5' guanosine-triphosphate cap, 5'N7-Methylguanosine-triphosphate cap, 5'triphosphate cap, 3'phosphate, 3'thiophosphate, 5'phosphate, 5'thiophosphate, Cis-Syn thymidine dimer, trimers, C12 spacer, C3 spacer, C6 spacer, dSpacer, PC spacer, rSpacer, Spacer 18, Spacer 9,3'-3' modifications, 5'-5' modifications, abasic, acridine, azobenzene, biotin, biotin BB, biotin TEG, cholesteryl TEG, desthiobiotin TEG, DNP TEG, DNP-X, DOTA, dT-Biotin, dual biotin, PC biotin, psoralen C2, psoralen C6, TINA, 3'DABCYL, black hole quencher 1, black hole quencer 2, DABCYL SE, dT-DABCYL, IRDye QC-1, QSY-21, QSY-35, QSY-7, QSY-9, carboxyl linker, thiol linkers, 2'-deoxyribonucleoside analog purine, 2'-deoxyribonucleoside analog pyrimidine, ribonucleoside analog, 2'-O-methyl ribonucleoside analog, sugar modified analogs, wobble/universal bases, fluorescent dye label, 2'-fluoro RNA, 2'-O-methyl RNA, methylphosphonate, phosphodiester DNA, phosphodiester RNA, phosphothioate DNA, phosphorothioate RNA, UNA, pseudouridine-5'-triphosphate, 5'-methylcytidine-5'-triphosphate, or any combination thereof. In some embodiments, a modification is one or more of 2'-*O*-methyl (2'-OMe), phosphorothioate (PS), 2'-*O*-methyl thioPACE (MSP), 2'-*O*-methyl-PACE (MP), 2'-fluoro RNA (2'-F-RNA), and constrained ethyl (S-cEt).

In some embodiments, a modification is permanent. In other embodiments, a modification is transient. In some embodiments, multiple modifications are made to a gRNA or a guide polynucleotide. A gRNA or a guide polynucleotide modification can alter physiochemical properties of a nucleotide, such as their conformation, polarity, hydrophobicity, chemical reactivity, base-pairing interactions, or any combination thereof.

A modification can also be a phosphorothioate substitute. In some embodiments, a natural phosphodiester bond can be susceptible to rapid degradation by cellular nucleases

and; a modification of internucleotide linkage using phosphorothioate (PS) bond substitutes can be more stable towards hydrolysis by cellular degradation. A modification can increase stability in a gRNA or a guide polynucleotide. A modification can also enhance biological activity. In some embodiments, a phosphorothioate enhanced RNA gRNA can inhibit RNase A, RNase T1, calf serum nucleases, or any combinations thereof. These properties can allow the use of PS-RNA gRNAs to be used in applications where exposure to nucleases is of high probability *in vivo* or *in vitro*. For example, phosphorothioate (PS) bonds can be introduced between the last 3-5 nucleotides at the 5'- or ''-end of a gRNA which can inhibit exonuclease degradation. In some embodiments, phosphorothioate bonds can be added throughout an entire gRNA to reduce attack by endonucleases. In some embodiments, the guide RNA is designed to disrupt a splice site (*i.e.*, a splice acceptor (SA) or a splice donor (SD). In some embodiments, the guide RNA is designed such that the base editing results in a premature STOP codon.

Protospacer Adjacent Motif

The term "protospacer adjacent motif (PAM)" or PAM-like motif refers to a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by the Cas9 nuclease in the CRISPR bacterial adaptive immune system. In some embodiments, the PAM can be a 5' PAM (*i.e.*, located upstream of the 5' end of the protospacer). In other embodiments, the PAM can be a 3' PAM (*i.e.*, located downstream of the 5' end of the protospacer). The PAM sequence is essential for target binding, but the exact sequence depends on a type of Cas protein. The PAM sequence can be any PAM sequence known in the art. Suitable PAM sequences include, but are not limited to, NGG, NGA, NGC, NGN, NGT, NGTT, NGCG, NGAG, NGAN, NGNG, NGCN, NGCG, NGTN, NNGRRT, NNNRRT, NNGRR(N), TTTV, TYCV, TYCV, TATV, NNNNGATT, NNAGAAW, or NAAAAC. Y is a pyrimidine; N is any nucleotide base; W is A or T.

A base editor provided herein can comprise a CRISPR protein-derived domain that is capable of binding a nucleotide sequence that contains a canonical or non-canonical protospacer adjacent motif (PAM) sequence. A PAM site is a nucleotide sequence in proximity to a target polynucleotide sequence. Some aspects of the disclosure provide for base editors comprising all or a portion of CRISPR proteins that have different PAM specificities.

For example, typically Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the G is guanine. A PAM can be CRISPR protein-specific and can be different between different base editors comprising different CRISPR protein-derived domains. A PAM can be 5' or 3' of a target sequence. A PAM can be upstream or downstream of a target sequence. A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. Often, a PAM is between 2-6 nucleotides in length.

In some embodiments, the PAM is an "NRN" PAM where the "N" in "NRN" is adenine (A), thymine (T), guanine (G), or cytosine (C), and the R is adenine (A) or guanine (G); or the PAM is an "NYN" PAM, wherein the "N" in NYN is adenine (A), thymine (T), guanine (G), or cytosine (C), and the Y is cytidine (C) or thymine (T), for example, as described in R.T. Walton *et al.*, 2020, *Science*, 10.1126/science.aba8853 (2020), the entire contents of which are incorporated herein by reference.

Several PAM variants are described in **Table 2** below.

Table 2. Cas9 proteins and corresponding PAM sequences

Variant	PAM
spCas9	NGG
spCas9-VRQR	NGA
spCas9-VRER	NGCG
xCas9 (sp)	NGN
saCas9	NNGRRT
saCas9-KKH	NNNRRT
spCas9-MQKSER	NGCG
spCas9-MQKSER	NGCN
spCas9-LRKIQK	NGTN
spCas9-LRVSQK	NGTN
spCas9-LRVSQL	NGTN
spCas9-MQKFRAER	NGC
Cpf1	5′ (TTTV)
SpyMac	5′-NAA-3′

In some embodiments, the PAM is NGC. In some embodiments, the NGC PAM is recognized by a Cas9 variant. In some embodiments, the NGC PAM variant includes one or more amino acid substitutions selected from D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E, and T1337R (collectively termed "MQKFRAER").

In some embodiments, the PAM is NGT. In some embodiments, the NGT PAM is recognized by a Cas9 variant. In some embodiments, the NGT PAM variant is generated through targeted mutations at one or more residues 1335, 1337, 1135, 1136, 1218, and/or 1219. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1219, 1335, 1337, 1218. In some embodiments, the NGT PAM variant is created through targeted mutations at one or more residues 1135, 1136, 1218, 1219, and 1335. In some embodiments, the NGT PAM variant is selected from the set of targeted mutations provided in **Tables 3A** and **3B** below.

Table 3A: NGT PAM Variant Mutations at residues 1219, 1335, 1337, 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R
9	L	L	T	
10	L	L	R	
11	L	L	Q	
12	L	L	L	
13	F	I	T	
14	F	I	R	
15	F	I	Q	
16	F	I	L	
17	F	G	C	
18	H	L	N	
19	F	G	C	A
20	H	L	N	V
21	L	A	W	
22	L	A	F	
23	L	A	Y	
24	I	A	W	
25	I	A	F	
26	I	A	Y	

Table 3B: NGT PAM Variant Mutations at residues 1135, 1136, 1218, 1219, and 1335

Variant	D1135L	S1136R	G1218S	E1219V	R1335Q
27	G				
28	V				
29	I				
30		A			
31		W			
32		H			
33		K			
34			K		
35			R		
36			Q		
37			T		
38			N		
39				I	
40				A	
41				N	
42				Q	
43				G	
44				L	
45				S	
46				T	
47					L
48					I
49					V
50					N
51					S
52					T
53					F
54					Y
55	N1286Q	I1331F			

In some embodiments, the NGT PAM variant is selected from variant 5, 7, 28, 31, or 36 in **Table 3A** and **Table 3B**. In some embodiments, the variants have improved NGT PAM recognition.

In some embodiments, the NGT PAM variants have mutations at residues 1219, 1335, 1337, and/or 1218. In some embodiments, the NGT PAM variant is selected with mutations for improved recognition from the variants provided in **Table 4** below.

Table 4: NGT PAM Variant Mutations at residues 1219, 1335, 1337, and 1218

Variant	E1219V	R1335Q	T1337	G1218
1	F	V	T	
2	F	V	R	
3	F	V	Q	
4	F	V	L	
5	F	V	T	R
6	F	V	R	R
7	F	V	Q	R
8	F	V	L	R

In some embodiments, the NGT PAM is selected from the variants provided in **Table 5** below.

Table 5. NGT PAM variants

	NGTN variant	D1135	S1136	G1218	E1219	A1322R	R1335	T1337
Variant 1	LRKIQK	L	R	K	I	-	Q	K
Variant 2	LRSVQK	L	R	S	V	-	Q	K
Variant 3	LRSVQL	L	R	S	V	-	Q	L
Variant 4	LRKIRQK	L	R	K	I	R	Q	K
Variant 5	LRSVRQK	L	R	S	V	R	Q	K
Variant 6	LRSVRQL	L	R	S	V	R	Q	L

In some embodiments the NGTN variant is variant 1. In some embodiments, the NGTN variant is variant 2. In some embodiments, the NGTN variant is variant 3. In some embodiments, the NGTN variant is variant 4. In some embodiments, the NGTN variant is variant 5. In some embodiments, the NGTN variant is variant 6.

In some embodiments, the Cas9 domain is a Cas9 domain from *Streptococcus pyogenes* (SpCas9). In some embodiments, the SpCas9 domain is a nuclease active SpCas9, a nuclease inactive SpCas9 (SpCas9d), or a SpCas9 nickase (SpCas9n). In some embodiments, the SpCas9 comprises a D9X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid except for D. In some embodiments, the SpCas9 comprises a D9A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having an NGG, a NGA, or a NGCG PAM sequence.

In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135E, R1335Q, and T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135E, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises one or more of a D1135X, a G1218X, a R1335X, and a T1337X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, the SpCas9 domain comprises one or more of a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the SpCas9 domain comprises a D1135V, a G1218R, a R1335Q, and a T1337R mutation, or corresponding mutations in any of the amino acid sequences provided herein.

In some examples, a PAM recognized by a CRISPR protein-derived domain of a base editor disclosed herein can be provided to a cell on a separate oligonucleotide to an insert (*e.g.*, an AAV insert) encoding the base editor. In such embodiments, providing PAM on a separate oligonucleotide can allow cleavage of a target sequence that otherwise would not be able to be cleaved, because no adjacent PAM is present on the same polynucleotide as the target sequence.

In an embodiment, *S. pyogenes* Cas9 (SpCas9) can be used as a CRISPR endonuclease for genome engineering. However, others can be used. In some embodiments, a different endonuclease can be used to target certain genomic targets. In some embodiments, synthetic SpCas9-derived variants with non-NGG PAM sequences can be used. Additionally, other Cas9 orthologues from various species have been identified and these "non-SpCas9s" can bind a variety of PAM sequences that can also be useful for the

present disclosure. For example, the relatively large size of SpCas9 (approximately 4kb coding sequence) can lead to plasmids carrying the SpCas9 cDNA that cannot be efficiently expressed in a cell. Conversely, the coding sequence for *Staphylococcus aureus* Cas9 (SaCas9) is approximately 1 kilobase shorter than SpCas9, possibly allowing it to be efficiently expressed in a cell. Similar to SpCas9, the SaCas9 endonuclease is capable of modifying target genes in mammalian cells *in vitro* and in mice *in vivo*. In some embodiments, a Cas protein can target a different PAM sequence. In some embodiments, a target gene can be adjacent to a Cas9 PAM, 5'-NGG, for example. In other embodiments, other Cas9 orthologs can have different PAM requirements. For example, other PAMs such as those of *S. thermophilus* (5'-NNAGAA for CRISPR1 and 5'-NGGNG for CRISPR3) and *Neisseria meningitidis* (5'-NNNNGATT) can also be found adjacent to a target gene.

In some embodiments, for a *S. pyogenes* system, a target gene sequence can precede (*i.e.*, be 5' to) a 5'-NGG PAM, and a 20-nt guide RNA sequence can base pair with an opposite strand to mediate a Cas9 cleavage adjacent to a PAM. In some embodiments, an adjacent cut can be or can be about 3 base pairs upstream of a PAM. In some embodiments, an adjacent cut can be or can be about 10 base pairs upstream of a PAM. In some embodiments, an adjacent cut can be or can be about 0-20 base pairs upstream of a PAM. For example, an adjacent cut can be next to, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 base pairs upstream of a PAM. An adjacent cut can also be downstream of a PAM by 1 to 30 base pairs.

In some embodiments, the Cas9 domain is a recombinant Cas9 domain. In some embodiments, the recombinant Cas9 domain is a SpyMacCas9 domain. In some embodiments, the SpyMacCas9 domain is a nuclease active SpyMacCas9, a nuclease inactive SpyMacCas9 (SpyMacCas9d), or a SpyMacCas9 nickase (SpyMacCas9n). In some embodiments, the SaCas9 domain, the SaCas9d domain, or the SaCas9n domain can bind to a nucleic acid sequence having a non-canonical PAM. In some embodiments, the SpyMacCas9 domain, the SpCas9d domain, or the SpCas9n domain can bind to a nucleic acid sequence having a NAA PAM sequence.

The sequence of an exemplary Cas9 A homolog of Spy Cas9 in Streptococcus macacae with native 5'-NAAN-3' PAM specificity is known in the art and described, for example, by Chatterjee, et al., "A Cas9 with PAM recognition for adenine dinucleotides", Nature Communications, vol. 11, article no. 2474 (2020), and is in the Sequence Listing as SEQ ID NO: 237.

The sequence of an exemplary Cas9 A homolog of Spy Cas9 in *Streptococcus macacae* with native 5'-NAAN-3' PAM specificity is known in the art and described, for example, by Jakimo *et al.* (biorxiv.org/content/biorxiv/early/2018/09/27/429654.full.pdf), In some embodiments, a variant Cas9 protein harbors, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a reduced ability to cleave a target DNA or RNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). As another non-limiting example, in some embodiments, the variant Cas9 protein harbors D10A, H840A, P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations such that the polypeptide has a reduced ability to cleave a target DNA. Such a Cas9 protein has a reduced ability to cleave a target DNA (*e.g.*, a single stranded target DNA) but retains the ability to bind a target DNA (*e.g.*, a single stranded target DNA). In some embodiments, when a variant Cas9 protein harbors W476A and W1126A mutations or when the variant Cas9 protein harbors P475A, W476A, N477A, D1125A, W1126A, and D1218A mutations, the variant Cas9 protein does not bind efficiently to a PAM sequence. Thus, in some such cases, when such a variant Cas9 protein is used in a method of binding, the method does not require a PAM sequence. In other words, in some embodiments, when such a variant Cas9 protein is used in a method of binding, the method can include a guide RNA, but the method can be performed in the absence of a PAM sequence (and the specificity of binding is therefore provided by the targeting segment of the guide RNA). Other residues can be mutated to achieve the above effects (*i.e.*, inactivate one or the other nuclease portions). As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 can be altered (*i.e.*, substituted). Also, mutations other than alanine substitutions are suitable.

In some embodiments, a CRISPR protein-derived domain of a base editor can comprise all or a portion of a Cas9 protein with a canonical PAM sequence (NGG). In other embodiments, a Cas9-derived domain of a base editor can employ a non-canonical PAM sequence. Such sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" Nature 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition" Nature Biotechnology 33, 1293-1298 (2015); R.T. Walton *et al.* "Unconstrained genome

targeting with near-PAMless engineered CRISPR-Cas9 variants" *Science* 10.1126/science.aba8853 (2020); Hu *et al.* "Evolved Cas9 variants with broad PAM compatibility and high DNA specificity," *Nature*, 2018 Apr. 5, 556(7699), 57-63; Miller *et al.*, "Continuous evolution of SpCas9 variants compatible with non-G PAMs" *Nat. Biotechnol.*, 2020 Apr;38(4):471-481; the entire contents of each are hereby incorporated by reference.

Cas9 Domains with Reduced Exclusivity

Typically, Cas9 proteins, such as Cas9 from *S. pyogenes* (spCas9), require a canonical NGG PAM sequence to bind a particular nucleic acid region, where the "N" in "NGG" is adenosine (A), thymidine (T), or cytosine (C), and the G is guanosine. This may limit the ability to edit desired bases within a genome. In some embodiments, the base editing fusion proteins provided herein may need to be placed at a precise location, for example a region comprising a target base that is upstream of the PAM. See *e.g.*, Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" *Nature* 533, 420-424 (2016), the entire contents of which are hereby incorporated by reference. Exemplary polypeptide sequences for spCas9 proteins capable of binding a PAM sequence are provided in the Sequence Listing as SEQ ID NOs: 197, 201, and 234-237. Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (*e.g.*, NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described in Kleinstiver, B. P., *et al.*, "Engineered CRISPR-Cas9 nucleases with altered PAM specificities" *Nature* 523, 481-485 (2015); and Kleinstiver, B. P., *et al.*, "Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition" *Nature Biotechnology* 33, 1293-1298 (2015); the entire contents of each are hereby incorporated by reference.

Fusion proteins with Internal Insertions

Provided herein are fusion proteins comprising a heterologous polypeptide fused to a nucleic acid programmable nucleic acid binding protein, for example, a napDNAbp. A heterologous polypeptide can be a polypeptide that is not found in the native or wild-type

napDNAbp polypeptide sequence. The heterologous polypeptide can be fused to the napDNAbp at a C-terminal end of the napDNAbp, an N-terminal end of the napDNAbp, or inserted at an internal location of the napDNAbp.

In some embodiments, the heterologous polypeptide is inserted at an internal location of the napDNAbp. In some embodiments, the heterologous polypeptide is a deaminase (*e.g.*, adenosine deaminase) or a functional fragment thereof. For example, a fusion protein can comprise a deaminase (*e.g.*, adenosine deaminase) flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 polypeptide. The deaminase in a fusion protein can be an adenosine deaminase. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10 or a variant thereof).

In some embodiments, the fusion protein comprises the structure:

NH2-[N-terminal fragment of a napDNAbp]-[deaminase]-[C-terminal fragment of a napDNAbp]-COOH;

NH2-[N-terminal fragment of a Cas9]-[adenosine deaminase]-[C-terminal fragment of a Cas9]-COOH;

wherein each instance of "]-[" is an optional linker.

The deaminase can be a circular permutant deaminase. For example, the deaminase can be a circular permutant adenosine deaminase. In some embodiments, the deaminase is a circular permutant TadA, circularly permutated at amino acid residue 116 as numbered in the TadA reference sequence. In some embodiments, the deaminase is a circular permutant TadA, circularly permutated at amino acid residue 136 as numbered in the TadA reference sequence. In some embodiments, the deaminase is a circular permutant TadA, circularly permutated at amino acid residue 65 as numbered in the TadA reference sequence.

The fusion protein can comprise more than one deaminase. The fusion protein can comprise, for example, 1, 2, 3, 4, 5 or more deaminases. In some embodiments, the fusion protein comprises one deaminase. In some embodiments, the fusion protein comprises two deaminases. The two or more deaminases in a fusion protein can be an adenosine deaminase, cytidine deaminase, or a combination thereof. The two or more deaminases can be homodimers. The two or more deaminases can be heterodimers. The two or more deaminases can be inserted in tandem in the napDNAbp. In some embodiments, the two or more deaminases may not be in tandem in the napDNAbp.

In some embodiments, the napDNAbp in the fusion protein is a Cas9 polypeptide or a fragment thereof. The Cas9 polypeptide can be a variant Cas9 polypeptide. In some

embodiments, the Cas9 polypeptide is a Cas9 nickase (nCas9) polypeptide or a fragment thereof. In some embodiments, the Cas9 polypeptide is a nuclease dead Cas9 (dCas9) polypeptide or a fragment thereof. The Cas9 polypeptide in a fusion protein can be a full-length Cas9 polypeptide. In some cases, the Cas9 polypeptide in a fusion protein may not be a full length Cas9 polypeptide. The Cas9 polypeptide can be truncated, for example, at a N-terminal or C-terminal end relative to a naturally-occurring Cas9 protein. The Cas9 polypeptide can be a circularly permuted Cas9 protein. The Cas9 polypeptide can be a fragment, a portion, or a domain of a Cas9 polypeptide, that is still capable of binding the target polynucleotide and a guide nucleic acid sequence.

In some embodiments, the Cas9 polypeptide is a *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus 1* Cas9 (St1Cas9), or fragments or variants thereof.

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-terminal fragments of a Cas9 polypeptide are also useful for base editing in the methods as described herein. Fusion proteins comprising Cas9 and one or more deaminase domains, *e.g.*, adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas9 sequences are also useful for highly specific and efficient base editing of target sequences. In an embodiment, a chimeric Cas9 fusion protein contains a heterologous catalytic domain (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) inserted within a Cas9 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within a Cas9. In some embodiments, an adenosine deaminase is fused within a Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and an adenosine deaminase fused to the N-terminus.

Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas9 are provided as follows:

NH_2-[Cas9(adenosine deaminase)]-[cytidine deaminase]-COOH;

NH_2-[cytidine deaminase]-[Cas9(adenosine deaminase)]-COOH;

NH_2-[Cas9(cytidine deaminase)]-[adenosine deaminase]-COOH; or

NH_2-[adenosine deaminase]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*, deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10). In some embodiments, the TadA is a TadA variant. In some embodiments, a TadA variant is fused within Cas9 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA variant is fused within Cas9 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA variant is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas9 and a TadA variant fused to the N-terminus. Exemplary structures of a fusion protein with a TadA variant and a cytidine deaminase and a Cas9 are provided as follows:

NH_2-[Cas9(TadA variant)]-[cytidine deaminase]-COOH;

NH_2-[cytidine deaminase]-[Cas9(TadA variant)]-COOH;

NH_2-[Cas9(cytidine deaminase)]-[TadA variant]-COOH; or

NH_2-[TadA variant]-[Cas9(cytidine deaminase)]-COOH.

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In other embodiments, the fusion protein contains one or more catalytic domains. In other embodiments, at least one of the one or more catalytic domains is inserted within the Cas12 polypeptide or is fused at the Cas12 N- terminus or C-terminus. In other embodiments, at least one of the one or more catalytic domains is inserted within a loop, an alpha helix region, an unstructured portion, or a solvent accessible portion of the Cas12 polypeptide. In other embodiments, the Cas12 polypeptide is Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the Cas12 polypeptide has at least about 85% amino acid sequence identity to Bacillus hisashii Cas12b, Bacillus thermoamylovorans Cas12b, Bacillus sp. V3-13 Cas12b, or Alicyclobacillus acidiphilus Cas12b (SEQ ID NO: 254). In other embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to Bacillus hisashii Cas12b (SEQ ID NO: 255), Bacillus thermoamylovorans Cas12b, Bacillus sp. V3-13 Cas12b, or Alicyclobacillus acidiphilus Cas12b. In other embodiments, the Cas12 polypeptide has at least about 95% amino acid sequence identity to Bacillus hisashii Cas12b, Bacillus thermoamylovorans

Cas12b (SEQ ID NO: 256), Bacillus sp. V3-13 Cas12b (SEQ ID NO: 257), or Alicyclobacillus acidiphilus Cas12b.

In other embodiments, the Cas12 polypeptide contains or consists essentially of a fragment of Bacillus hisashii Cas12b, Bacillus thermoamylovorans Cas12b, Bacillus sp. V3-13 Cas12b, or Alicyclobacillus acidiphilus Cas12b. In embodiments, the Cas12 polypeptide contains BvCas12b (V4), which in some embodiments is expressed as 5' mRNA Cap---5' UTR---bhCas12b---STOP sequence --- 3' UTR --- 120polyA tail (SEQ ID NOs: 258-260).

In other embodiments, the catalytic domain is inserted between amino acid positions 153-154, 255-256, 306-307, 980-981, 1019-1020, 534-535, 604-605, or 344-345 of BhCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P153 and S154 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K255 and E256 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids D980 and G981 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1019 and L1020 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids F534 and P535 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K604 and G605 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids H344 and F345 of BhCas12b. In other embodiments, catalytic domain is inserted between amino acid positions 147 and 148, 248 and 249, 299 and 300, 991 and 992, or 1031 and 1032 of BvCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P147 and D148 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G248 and G249 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids P299 and E300 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G991 and E992 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1031 and M1032 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acid positions 157 and 158, 258 and 259, 310 and 311, 1008 and 1009, or 1044 and 1045 of AaCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P157 and G158 of AaCas12b. In other

embodiments, the catalytic domain is inserted between amino acids V258 and G259 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids D310 and P311 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1008 and E1009 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1044 and K1045 at of AaCas12b.

In other embodiments, the fusion protein contains a nuclear localization signal (e.g., a bipartite nuclear localization signal). In other embodiments, the amino acid sequence of the nuclear localization signal is MAPKKKRKVGIHGVPAA (SEQ ID NO: 261). In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence:

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC (SEQ ID NO: 262). In other embodiments, the Cas12b polypeptide contains a mutation that silences the catalytic activity of a RuvC domain. In other embodiments, the Cas12b polypeptide contains D574A, D829A and/or D952A mutations. In other embodiments, the fusion protein further contains a tag (e.g., an influenza hemagglutinin tag).

In some embodiments, the fusion protein comprises a napDNAbp domain (e.g., Cas12-derived domain) with an internally fused nucleobase editing domain (e.g., all or a portion of a deaminase domain, e.g., an adenosine deaminase domain). In some embodiments, the napDNAbp is a Cas12b.

By way of nonlimiting example, an adenosine deaminase (e.g., TadA*8.13) may be inserted into a BhCas12b to produce a fusion protein (e.g., TadA*8.13-BhCas12b) that effectively edits a nucleic acid sequence.

In some embodiments, the base editing system described herein is an ABE with TadA inserted into a Cas9. Polypeptide sSequences of relevant ABEs with TadA inserted into a Cas9 are provided in the attached Ssequence Llisting as SEQ ID NOs: 263-308.

In some embodiments, adenosine base editors were generated to insert TadA or variants thereof into the Cas9 polypeptide at the identified positions.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/US2020/016285 and U.S. Provisional Application Nos. 62/852,228 and 62/852,224, the contents of which are incorporated by reference herein in their entireties.

The heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp (*e.g.*, Cas9) at a suitable location, for example, such that the napDNAbp retains its ability to bind the target polynucleotide and a guide nucleic acid. A deaminase (*e.g.*, adenosine

deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted into a napDNAbp without compromising function of the deaminase (*e.g.*, base editing activity) or the napDNAbp (*e.g.*, ability to bind to target nucleic acid and guide nucleic acid). A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp at, for example, a disordered region or a region comprising a high temperature factor or B-factor as shown by crystallographic studies. Regions of a protein that are less ordered, disordered, or unstructured, for example solvent exposed regions and loops, can be used for insertion without compromising structure or function. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted in the napDNAbp in a flexible loop region or a solvent-exposed region. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in a flexible loop of the Cas9 polypeptide.

In some embodiments, the insertion location of a deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is determined by B-factor analysis of the crystal structure of Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted in regions of the Cas9 polypeptide comprising higher than average B-factors (*e.g.*, higher B factors compared to the total protein or the protein domain comprising the disordered region). B-factor or temperature factor can indicate the fluctuation of atoms from their average position (for example, as a result of temperature-dependent atomic vibrations or static disorder in a crystal lattice). A high B-factor (*e.g.*, higher than average B-factor) for backbone atoms can be indicative of a region with relatively high local mobility. Such a region can be used for inserting a deaminase without compromising structure or function. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a Cα atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200%, or greater than 200% more than the average B-factor for the total protein. A deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at a location with a residue having a Cα atom with a B-factor that is 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190%, 200% or greater than 200% more than the average B-factor for a Cas9 protein domain comprising the

residue. Cas9 polypeptide positions comprising a higher than average B-factor can include, for example, residues 768, 792, 1052, 1015, 1022, 1026, 1029, 1067, 1040, 1054, 1068, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence. Cas9 polypeptide regions comprising a higher than average B-factor can include, for example, residues 792-872, 792-906, and 2-791 as numbered in the above Cas9 reference sequence.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 791-792, 792-793, 1015-1016, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1052-1053, 1054-1055, 1067-1068, 1068-1069, 1247-1248, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 792-793, 793-794, 1016-1017, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1053-1054, 1055-1056, 1068-1069, 1069-1070, 1248-1249, or 1249-1250 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 791, 792, 1015, 1016, 1022, 1023, 1026, 1029, 1040, 1052, 1054, 1067, 1068, 1069, 1246, 1247, and 1248 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. It should be understood that the reference to the above Cas9 reference sequence with respect to insertion positions is for illustrative purposes. The insertions as discussed herein are not limited to the Cas9 polypeptide sequence of the above Cas9 reference sequence, but include insertion at corresponding locations in variant Cas9 polypeptides, for example a Cas9 nickase (nCas9), nuclease dead Cas9 (dCas9), a Cas9 variant lacking a nuclease domain, a truncated Cas9, or a Cas9 domain lacking partial or complete HNH domain.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 768-769, 792-793, 1022-1023, 1026-1027, 1029-1030, 1040-1041, 1068-1069, or 1247-1248 as numbered in the above Cas9 reference sequence or

corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide is inserted between amino acid positions 769-770, 793-794, 1023-1024, 1027-1028, 1030-1031, 1041-1042, 1069-1070, or 1248-1249 as numbered in the above Cas9 reference sequence or corresponding amino acid positions thereof. In some embodiments, the heterologous polypeptide replaces an amino acid residue selected from the group consisting of: 768, 792, 1022, 1026, 1040, 1068, and 1247 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue as described herein, or a corresponding amino acid residue in another Cas9 polypeptide. In an embodiment, a heterologous polypeptide (*e.g.*, deaminase) can be inserted in the napDNAbp at an amino acid residue selected from the group consisting of: 1002, 1003, 1025, 1052-1056, 1242-1247, 1061-1077, 943-947, 686-691, 569-578, 530-539, and 1060-1077 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) can be inserted at the N-terminus or the C-terminus of the residue or replace the residue. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of the residue.

In some embodiments, an adenosine deaminase (*e.g.*, TadA) is inserted at an amino acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, an adenosine deaminase (*e.g.*, TadA) is inserted in place of residues 792-872, 792-906, or 2-791 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the N-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted at the C-terminus of an amino acid selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the adenosine deaminase is inserted to replace an amino acid selected

from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the C-terminus of amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted to replace amino acid residue 768 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 791 or is inserted at amino acid residue 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the N-terminus of amino acid residue 791 or is inserted at the N-terminus of amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid 791 or is inserted at the N-terminus of amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid 791, or is inserted to replace amino acid 792, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*,

adenosine deaminase) is inserted at the N-terminus of amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1016 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1022, or is inserted at amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the N-terminus of amino acid residue 1022 or is inserted at the N-terminus of amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid residue 1022 or is inserted at the C-terminus of amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1022, or is inserted to replace amino acid residue 1023, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1026, or is inserted at amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the N-terminus of amino acid residue 1026 or is inserted at the N-terminus of amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid residue 1026 or is inserted at the C-terminus of amino acid residue 1029, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1026, or is

inserted to replace amino acid residue 1029, as numbered in the above Cas9 reference sequence, or corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase,) is inserted at the N-terminus of amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1040 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1052, or is inserted at amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) is inserted at the N-terminus of amino acid residue 1052 or is inserted at the N-terminus of amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid residue 1052 or is inserted at the C-terminus of amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1052, or is inserted to replace amino acid residue 1054, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1067, or is inserted at amino acid residue 1068, or is inserted at amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the N-terminus of amino acid residue 1067 or is inserted at the N-terminus of amino acid residue 1068 or is inserted at the N-terminus of amino acid

residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the C-terminus of amino acid residue 1067 or is inserted at the C-terminus of amino acid residue 1068 or is inserted at the C-terminus of amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1067, or is inserted to replace amino acid residue 1068, or is inserted to replace amino acid residue 1069, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at amino acid residue 1246, or is inserted at amino acid residue 1247, or is inserted at amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted at the N-terminus of amino acid residue 1246 or is inserted at the N-terminus of amino acid residue 1247 or is inserted at the N-terminus of amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase is inserted at the C-terminus of amino acid residue 1246 or is inserted at the C-terminus of amino acid residue 1247 or is inserted at the C-terminus of amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deaminase (*e.g.*, adenosine deaminase) is inserted to replace amino acid residue 1246, or is inserted to replace amino acid residue 1247, or is inserted to replace amino acid residue 1248, as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

In some embodiments, a heterologous polypeptide (*e.g.*, deaminase) is inserted in a flexible loop of a Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of 530-537, 569-570, 686-691, 943-947, 1002-1025, 1052-1077, 1232-1247, or 1298-1300 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The flexible loop portions can be selected from the group consisting of: 1-529, 538-568, 580-685, 692-942, 948-1001, 1026-1051, 1078-1231, or

1248-1297 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, adenine deaminase) can be inserted into a Cas9 polypeptide region corresponding to amino acid residues: 1017-1069, 1242-1247, 1052-1056, 1060-1077, 1002 – 1003, 943-947, 530-537, 568-579, 686-691, 1242-1247, 1298 – 1300, 1066-1077, 1052-1056, or 1060-1077 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

A heterologous polypeptide (*e.g.*, adenine deaminase) can be inserted in place of a deleted region of a Cas9 polypeptide. The deleted region can correspond to an N-terminal or C-terminal portion of the Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-872 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 792-906 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 2-791 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. In some embodiments, the deleted region corresponds to residues 1017-1069 as numbered in the above Cas9 reference sequence, or corresponding amino acid residues thereof.

Exemplary internal fusions base editors are provided in **Table 6** below:

Table 6: Insertion loci in Cas9 proteins

BE ID	Modification	Other ID
IBE001	Cas9 TadA ins 1015	ISLAY01
IBE002	Cas9 TadA ins 1022	ISLAY02
IBE003	Cas9 TadA ins 1029	ISLAY03
IBE004	Cas9 TadA ins 1040	ISLAY04
IBE005	Cas9 TadA ins 1068	ISLAY05
IBE006	Cas9 TadA ins 1247	ISLAY06
IBE007	Cas9 TadA ins 1054	ISLAY07
IBE008	Cas9 TadA ins 1026	ISLAY08
IBE009	Cas9 TadA ins 768	ISLAY09
IBE020	delta HNH TadA 792	ISLAY20
IBE021	N-term fusion single TadA helix truncated 165-end	ISLAY21

IBE029	TadA-Circular Permutant116 ins1067	ISLAY29
IBE031	TadA- Circular Permutant 136 ins1248	ISLAY31
IBE032	TadA- Circular Permutant 136ins 1052	ISLAY32
IBE035	delta 792-872 TadA ins	ISLAY35
IBE036	delta 792-906 TadA ins	ISLAY36
IBE043	TadA-Circular Permutant 65 ins1246	ISLAY43
IBE044	TadA ins C-term truncate2 791	ISLAY44

A heterologous polypeptide (*e.g.*, deaminase) can be inserted within a structural or functional domain of a Cas9 polypeptide. A heterologous polypeptide (*e.g.*, deaminase) can be inserted between two structural or functional domains of a Cas9 polypeptide. A heterologous polypeptide (*e.g.*, deaminase) can be inserted in place of a structural or functional domain of a Cas9 polypeptide, for example, after deleting the domain from the Cas9 polypeptide. The structural or functional domains of a Cas9 polypeptide can include, for example, RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH.

In some embodiments, the Cas9 polypeptide lacks one or more domains selected from the group consisting of: RuvC I, RuvC II, RuvC III, Rec1, Rec2, PI, or HNH domain. In some embodiments, the Cas9 polypeptide lacks a nuclease domain. In some embodiments, the Cas9 polypeptide lacks an HNH domain. In some embodiments, the Cas9 polypeptide lacks a portion of the HNH domain such that the Cas9 polypeptide has reduced or abolished HNH activity. In some embodiments, the Cas9 polypeptide comprises a deletion of the nuclease domain, and the deaminase is inserted to replace the nuclease domain. In some embodiments, the HNH domain is deleted and the deaminase is inserted in its place. In some embodiments, one or more of the RuvC domains is deleted and the deaminase is inserted in its place.

A fusion protein comprising a heterologous polypeptide can be flanked by a N-terminal and a C-terminal fragment of a napDNAbp. In some embodiments, the fusion protein comprises a deaminase flanked by a N- terminal fragment and a C-terminal fragment of a Cas9 polypeptide. The N terminal fragment or the C terminal fragment can bind the target polynucleotide sequence. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of a flexible loop of a Cas9 polypeptide. The C-terminus of the N terminal fragment or the N-terminus of the C terminal fragment can comprise a part of an alpha-helix structure of the Cas9 polypeptide. The N-

terminal fragment or the C-terminal fragment can comprise a DNA binding domain. The N-terminal fragment or the C-terminal fragment can comprise a RuvC domain. The N-terminal fragment or the C-terminal fragment can comprise an HNH domain. In some embodiments, neither of the N-terminal fragment and the C-terminal fragment comprises an HNH domain.

In some embodiments, the C-terminus of the N terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. In some embodiments, the N-terminus of the C terminal Cas9 fragment comprises an amino acid that is in proximity to a target nucleobase when the fusion protein deaminates the target nucleobase. The insertion location of different deaminases can be different in order to have proximity between the target nucleobase and an amino acid in the C-terminus of the N terminal Cas9 fragment or the N-terminus of the C terminal Cas9 fragment. For example, the insertion position of a deaminase can be at an amino acid residue selected from the group consisting of: 1015, 1022, 1029, 1040, 1068, 1247, 1054, 1026, 768, 1067, 1248, 1052, and 1246 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment of a fusion protein (*i.e.* the N-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the N-terminus of a Cas9 polypeptide. The N-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino acids. The N-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1-56, 1-95, 1-200, 1-300, 1-400, 1-500, 1-600, 1-700, 1-718, 1-765, 1-780, 1-906, 1-918, or 1-1100 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The C-terminal Cas9 fragment of a fusion protein (*i.e.* the C-terminal Cas9 fragment flanking the deaminase in a fusion protein) can comprise the C-terminus of a Cas9 polypeptide. The C-terminal Cas9 fragment of a fusion protein can comprise a length of at least about: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, or 1300 amino

acids. The C-terminal Cas9 fragment of a fusion protein can comprise a sequence corresponding to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide. The N-terminal Cas9 fragment can comprise a sequence comprising at least: 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% sequence identity to amino acid residues: 1099-1368, 918-1368, 906-1368, 780-1368, 765-1368, 718-1368, 94-1368, or 56-1368 as numbered in the above Cas9 reference sequence, or a corresponding amino acid residue in another Cas9 polypeptide.

The N-terminal Cas9 fragment and C-terminal Cas9 fragment of a fusion protein taken together may not correspond to a full-length naturally occurring Cas9 polypeptide sequence, for example, as set forth in the above Cas9 reference sequence.

The fusion protein described herein can effect targeted deamination with reduced deamination at non-target sites (*e.g.*, off-target sites), such as reduced genome wide spurious deamination. The fusion protein described herein can effect targeted deamination with reduced bystander deamination at non-target sites. The undesired deamination or off-target deamination can be reduced by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide. The undesired deamination or off-target deamination can be reduced by at least one-fold, at least two-fold, at least three-fold, at least four-fold, at least five-fold, at least tenfold, at least fifteen fold, at least twenty fold, at least thirty fold, at least forty fold, at least fifty fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, or at least hundred fold, compared with, for example, an end terminus fusion protein comprising the deaminase fused to a N terminus or a C terminus of a Cas9 polypeptide.

In some embodiments, the deaminase (*e.g.*, adenosine deaminase) of the fusion protein deaminates no more than two nucleobases within the range of an R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than three nucleobases within the range of the R-loop. In some embodiments, the deaminase of the fusion protein deaminates no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleobases within the range of the R-loop. An R-loop is a three-stranded nucleic acid structure including a DNA:RNA hybrid, a DNA:DNA or an RNA: RNA complementary structure and the associated with single-stranded DNA. As used herein, an R-loop may be formed when a target polynucleotide is

contacted with a CRISPR complex or a base editing complex, wherein a portion of a guide polynucleotide, *e.g.* a guide RNA, hybridizes with and displaces with a portion of a target polynucleotide, *e.g.* a target DNA. In some embodiments, an R-loop comprises a hybridized region of a spacer sequence and a target DNA complementary sequence. An R-loop region may be of about 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, or 50 nucleobase pairs in length. In some embodiments, the R-loop region is about 20 nucleobase pairs in length. It should be understood that, as used herein, an R-loop region is not limited to the target DNA strand that hybridizes with the guide polynucleotide. For example, editing of a target nucleobase within an R-loop region may be to a DNA strand that comprises the complementary strand to a guide RNA, or may be to a DNA strand that is the opposing strand of the strand complementary to the guide RNA. In some embodiments, editing in the region of the R-loop comprises editing a nucleobase on non-complementary strand (protospacer strand) to a guide RNA in a target DNA sequence.

The fusion protein described herein can effect target deamination in an editing window different from canonical base editing. In some embodiments, a target nucleobase is from about 1 to about 20 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 2 to about 12 bases upstream of a PAM sequence in the target polynucleotide sequence. In some embodiments, a target nucleobase is from about 1 to 9 base pairs, about 2 to 10 base pairs, about 3 to 11 base pairs, about 4 to 12 base pairs, about 5 to 13 base pairs, about 6 to 14 base pairs, about 7 to 15 base pairs, about 8 to 16 base pairs, about 9 to 17 base pairs, about 10 to 18 base pairs, about 11 to 19 base pairs, about 12 to 20 base pairs, about 1 to 7 base pairs, about 2 to 8 base pairs, about 3 to 9 base pairs, about 4 to 10 base pairs, about 5 to 11 base pairs, about 6 to 12 base pairs, about 7 to 13 base pairs, about 8 to 14 base pairs, about 9 to 15 base pairs, about 10 to 16 base pairs, about 11 to 17 base pairs, about 12 to 18 base pairs, about 13 to 19 base pairs, about 14 to 20 base pairs, about 1 to 5 base pairs, about 2 to 6 base pairs, about 3 to 7 base pairs, about 4 to 8 base pairs, about 5 to 9 base pairs, about 6 to 10 base pairs, about 7 to 11 base pairs, about 8 to 12 base pairs, about 9 to 13 base pairs, about 10 to 14 base pairs, about 11 to 15 base pairs, about 12 to 16 base pairs, about 13 to 17 base pairs, about 14 to 18 base pairs, about 15 to 19 base pairs, about 16 to 20 base pairs, about 1 to 3 base pairs, about 2 to 4 base pairs, about 3 to 5 base pairs, about 4 to 6 base pairs, about 5 to 7 base pairs, about 6 to 8 base pairs, about 7 to 9 base pairs, about 8 to 10 base pairs, about 9 to 11 base pairs, about

10 to 12 base pairs, about 11 to 13 base pairs, about 12 to 14 base pairs, about 13 to 15 base pairs, about 14 to 16 base pairs, about 15 to 17 base pairs, about 16 to 18 base pairs, about 17 to 19 base pairs, about 18 to 20 base pairs away or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more base pairs away from or upstream of the PAM sequence. In some embodiments, a target nucleobase is about 1, 2, 3, 4, 5, 6, 7, 8, or 9 base pairs upstream of the PAM sequence. In some embodiments, a target nucleobase is about 2, 3, 4, or 6 base pairs upstream of the PAM sequence.

The fusion protein can comprise more than one heterologous polypeptide. For example, the fusion protein can additionally comprise one or more UGI domains and/or one or more nuclear localization signals. The two or more heterologous domains can be inserted in tandem. The two or more heterologous domains can be inserted at locations such that they are not in tandem in the NapDNAbp.

A fusion protein can comprise a linker between the deaminase and the napDNAbp polypeptide. The linker can be a peptide or a non-peptide linker. For example, the linker can be an XTEN, (GGGS)n (SEQ ID NO: 246), (GGGGS)n (SEQ ID NO: 247), (G)n, (EAAAK)n (SEQ ID NO: 248), (GGS)n, SGSETPGTSESATPES (SEQ ID NO: 249).

In some embodiments, the napDNAbp in the fusion protein is a Cas12 polypeptide, e.g., Cas12b/C2c1, or a fragment thereof. The Cas12 polypeptide can be a variant Cas12 polypeptide. In other embodiments, the N- or C-terminal fragments of the Cas12 polypeptide comprise a nucleic acid programmable DNA binding domain or a RuvC domain. In other embodiments, the fusion protein contains a linker between the Cas12 polypeptide and the catalytic domain. In other embodiments, the amino acid sequence of the linker is GGSGGS (SEQ ID NO: 250) or GSSGSETPGTSESATPESSG (SEQ ID NO: 251). In other embodiments, the linker is a rigid linker. In other embodiments of the above aspects, the linker is encoded by GGAGGCTCTGGAGGAAGC (SEQ ID NO: 252) or GGCTCTTCTGGATCTGAAACACCTGGCACAAGCGAGAGCGCCACCCCTGAGAGCTCTGGC (SEQ ID NO: 253).

In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the N-terminal and C-terminal fragments of napDNAbp are connected to the deaminase with a linker. In some embodiments, the N-terminal and C-terminal fragments are joined to the

deaminase domain without a linker. In some embodiments, the fusion protein comprises a linker between the N-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the C-terminal Cas9 fragment and the deaminase. In some embodiments, the fusion protein comprises a linker between the C-terminal Cas9 fragment and the deaminase, but does not comprise a linker between the N-terminal Cas9 fragment and the deaminase.

In some embodiments, the base editing system described herein is an ABE with TadA inserted into a Cas9. Sequences of relevant ABEs with TadA inserted into a Cas9 are provided. In some embodiments, adenosine deaminase base editors were generated to insert TadA or variants thereof into the Cas9 polypeptide at the identified positions.

Fusion proteins comprising a heterologous catalytic domain flanked by N- and C-terminal fragments of a Cas12 polypeptide are also useful for base editing in the methods as described herein. Fusion proteins comprising Cas12 and one or more deaminase domains, *e.g.*, adenosine deaminase, or comprising an adenosine deaminase domain flanked by Cas12 sequences are also useful for highly specific and efficient base editing of target sequences. In an embodiment, a chimeric Cas12 fusion protein contains a heterologous catalytic domain (*e.g.*, adenosine deaminase, cytidine deaminase, or adenosine deaminase and cytidine deaminase) inserted within a Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase domain and a cytidine deaminase domain inserted within a Cas12. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, an adenosine deaminase is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and an adenosine deaminase fused to the N-terminus. Exemplary structures of a fusion protein with an adenosine deaminase and a cytidine deaminase and a Cas12 are provided as follows:

NH2-[Cas12(adenosine deaminase)]-[cytidine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas12(adenosine deaminase)]-COOH;

NH2-[Cas12(cytidine deaminase)]-[adenosine deaminase]-COOH; or

NH2-[adenosine deaminase]-[Cas12(cytidine deaminase)]-COOH;

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In various embodiments, the catalytic domain has DNA modifying activity (*e.g.*, deaminase activity), such as adenosine deaminase activity. In some embodiments, the adenosine deaminase is a TadA (*e.g.*, TadA*7.10). In some embodiments, the TadA is a TadA*8. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase is fused to the C-terminus. In some embodiments, a TadA*8 is fused within Cas12 and a cytidine deaminase fused to the N-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 is fused to the C-terminus. In some embodiments, a cytidine deaminase is fused within Cas12 and a TadA*8 fused to the N-terminus. Exemplary structures of a fusion protein with a TadA*8 and a cytidine deaminase and a Cas12 are provided as follows:

N-[Cas12(TadA*8)]-[cytidine deaminase]-C;

N-[cytidine deaminase]-[Cas12(TadA*8)]-C;

N-[Cas12(cytidine deaminase)]-[TadA*8]-C; or

N-[TadA*8]-[Cas12(cytidine deaminase)]-C.

In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

In other embodiments, the fusion protein contains one or more catalytic domains. In other embodiments, at least one of the one or more catalytic domains is inserted within the Cas12 polypeptide or is fused at the Cas12 N- terminus or C-terminus. In other embodiments, at least one of the one or more catalytic domains is inserted within a loop, an alpha helix region, an unstructured portion, or a solvent accessible portion of the Cas12 polypeptide. In other embodiments, the Cas12 polypeptide is Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the Cas12 polypeptide has at least about 85% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b (SEQ ID NO: 254). In other embodiments, the Cas12 polypeptide has at least about 90% amino acid sequence identity to *Bacillus hisashii* Cas12b (SEQ ID NO: 255), *Bacillus thermoamylovorans* Cas12b, Bacillus sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide has at least about 95% amino acid sequence identity to *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans* Cas12b (SEQ ID NO: 256), *Bacillus* sp. V3-13 Cas12b (SEQ ID NO: 257), or *Alicyclobacillus acidiphilus* Cas12b. In other embodiments, the Cas12 polypeptide contains or consists essentially of a fragment of *Bacillus hisashii* Cas12b, *Bacillus thermoamylovorans*

Cas12b, *Bacillus* sp. V3-13 Cas12b, or *Alicyclobacillus acidiphilus* Cas12b. In embodiments, the Cas12 polypeptide contains BvCas12b (V4), which in some embodiments is expressed as 5' mRNA Cap---5' UTR---bhCas12b---STOP sequence --- 3' UTR --- 120polyA tail (SEQ ID NOs: 258-260).

In other embodiments, the catalytic domain is inserted between amino acid positions 153-154, 255-256, 306-307, 980-981, 1019-1020, 534-535, 604-605, or 344-345 of BhCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P153 and S154 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K255 and E256 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids D980 and G981 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1019 and L1020 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids F534 and P535 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids K604 and G605 of BhCas12b. In other embodiments, the catalytic domain is inserted between amino acids H344 and F345 of BhCas12b. In other embodiments, catalytic domain is inserted between amino acid positions 147 and 148, 248 and 249, 299 and 300, 991 and 992, or 1031 and 1032 of BvCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P147 and D148 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G248 and G249 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids P299 and E300 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids G991 and E992 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acids K1031 and M1032 of BvCas12b. In other embodiments, the catalytic domain is inserted between amino acid positions 157 and 158, 258 and 259, 310 and 311, 1008 and 1009, or 1044 and 1045 of AaCas12b or a corresponding amino acid residue of Cas12a, Cas12c, Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, or Cas12j/CasΦ. In other embodiments, the catalytic domain is inserted between amino acids P157 and G158 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids V258 and G259 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids D310 and P311 of AaCas12b. In other embodiments, the catalytic domain is inserted

between amino acids G1008 and E1009 of AaCas12b. In other embodiments, the catalytic domain is inserted between amino acids G1044 and K1045 at of AaCas12b.

In other embodiments, the fusion protein contains a nuclear localization signal (*e.g.*, a bipartite nuclear localization signal). In other embodiments, the amino acid sequence of the nuclear localization signal is MAPKKKRKVGIHGVPAA (SEQ ID NO: 261). In other embodiments of the above aspects, the nuclear localization signal is encoded by the following sequence:

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC (SEQ ID NO: 262). In other embodiments, the Cas12b polypeptide contains a mutation that silences the catalytic activity of a RuvC domain. In other embodiments, the Cas12b polypeptide contains D574A, D829A and/or D952A mutations. In other embodiments, the fusion protein further contains a tag (*e.g.*, an influenza hemagglutinin tag).

In some embodiments, the fusion protein comprises a napDNAbp domain (*e.g.*, Cas12-derived domain) with an internally fused nucleobase editing domain (*e.g.*, all or a portion of a deaminase domain, *e.g.*, an adenosine deaminase domain). In some embodiments, the napDNAbp is a Cas12b. In some embodiments, the base editor comprises a BhCas12b domain with an internally fused TadA*8 domain inserted at the loci provided in **Table 7** below.

Table 7: Insertion loci in Cas12b proteins

BhCas12b	Insertion site	Inserted between aa
position 1	153	PS
position 2	255	KE
position 3	306	DE
position 4	980	DG
position 5	1019	KL
position 6	534	FP
position 7	604	KG
position 8	344	HF
BvCas12b	Insertion site	Inserted between aa
position 1	147	PD
position 2	248	GG

position 3	299	PE
position 4	991	GE
position 5	1031	KM
AaCas12b	Insertion site	Inserted between aa
position 1	157	PG
position 2	258	VG
position 3	310	DP
position 4	1008	GE
position 5	1044	GK

By way of nonlimiting example, an adenosine deaminase (*e.g.*, TadA*8.13) may be inserted into a BhCas12b to produce a fusion protein (*e.g.*, TadA*8.13-BhCas12b) that effectively edits a nucleic acid sequence.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/US2020/016285, PCT/US2020/018073, PCT/US2020/018107, PCT/US2020/018124, PCT/US2020/018132, PCT/US2020/018169, PCT/US2020/018178, PCT/US2020/018192, PCT/US2020/018193, and PCT/US2020/018195, the contents of which are incorporated by reference herein in their entireties.

A to G Editing

In some embodiments, a base editor described herein comprises an adenosine deaminase domain. Such an adenosine deaminase domain of a base editor can facilitate the editing of an adenine (A) nucleobase to a guanine (G) nucleobase by deaminating the A to form inosine (I), which exhibits base pairing properties of G. Adenosine deaminase is capable of deaminating (i.e., removing an amine group) adenine of a deoxyadenosine residue in deoxyribonucleic acid (DNA). In some embodiments, an A-to-G base editor further comprises an inhibitor of inosine base excision repair, for example, a uracil glycosylase inhibitor (UGI) domain or a catalytically inactive inosine specific nuclease. Without wishing to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a deaminated adenosine residue (e.g., inosine), which can improve the activity or efficiency of the base editor.

A base editor comprising an adenosine deaminase can act on any polynucleotide, including DNA, RNA and DNA-RNA hybrids. In certain embodiments, a base editor comprising an adenosine deaminase can deaminate a target A of a polynucleotide comprising RNA. For example, the base editor can comprise an adenosine deaminase domain capable of deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on RNA (ADAR, e.g., ADAR1 or ADAR2) or tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion of an ADAT comprising one or more mutations which permit the ADAT to deaminate a target A in DNA. For example, the base editor can comprise all or a portion of an ADAT from Escherichia coli (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y, E155V, L84F, H123Y, I156F, or a corresponding mutation in another adenosine deaminase. Exemplary ADAT homolog polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 1 and 309-315.

In some embodiments, a base editor described herein comprises a fusion protein comprising an adenosine deaminase domain (*e.g.*, adenosine deaminase variant domain). In some embodiments, an adenosine deaminase variant domain contains a combination of alterations in a TadA*7.10 amino acid sequence, where the combinations are V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N. In some embodiments, the combinations of alterations in a TadA*7.10 amino acid sequence are V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; or L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N or a corresponding alteration in another adenosine deaminase.

Such an adenosine deaminase domain of a base editor can facilitate the editing of an adenine (A) nucleobase to a guanine (G) nucleobase by deaminating the A to form inosine (I), which exhibits base pairing properties of G. Adenosine deaminase is capable of deaminating (*i.e.*, removing an amine group) adenine of a deoxyadenosine residue in deoxyribonucleic acid (DNA).

In some embodiments, the nucleobase editors provided herein can be made by fusing together one or more protein domains, thereby generating a fusion protein. In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity (*e.g.*, efficiency, selectivity, and specificity) of the fusion proteins. For example, the fusion proteins provided herein can comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, the fusion proteins provided herein can have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9). Without wishing to be bound by any particular theory, the presence of the catalytic residue (*e.g.*, H840) maintains the activity of the Cas9 to cleave the non-edited (*e.g.*, non-deaminated) strand containing a T opposite the targeted A. Mutation of the catalytic residue (*e.g.*, D10 to A10) of Cas9 prevents cleavage of the edited strand containing the targeted A residue. Such Cas9 variants are able to generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a T to C change on the non-edited strand. In some embodiments, an A-to-G base editor further comprises an inhibitor of inosine base excision repair, for example, a uracil glycosylase inhibitor (UGI) domain or a catalytically inactive inosine specific nuclease. Without wishing to be bound by any particular theory, the UGI domain or catalytically inactive inosine specific nuclease can inhibit or prevent base excision repair of a deaminated adenosine residue (*e.g.*, inosine), which can improve the activity or efficiency of the base editor.

A base editor comprising an adenosine deaminase can act on any polynucleotide, including DNA, RNA and DNA-RNA hybrids. In certain embodiments, a base editor comprising an adenosine deaminase can deaminate a target A of a polynucleotide comprising RNA. For example, the base editor can comprise an adenosine deaminase domain capable of deaminating a target A of an RNA polynucleotide and/or a DNA-RNA hybrid polynucleotide. In an embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on RNA (ADAR, *e.g.*, ADAR1 or ADAR2). In another embodiment, an adenosine deaminase incorporated into a base editor comprises all or a portion of adenosine deaminase acting on tRNA (ADAT). A base editor comprising an adenosine deaminase domain can also be capable of deaminating an A nucleobase of a DNA polynucleotide. In an embodiment an adenosine deaminase domain of a base editor comprises all or a portion of an ADAT comprising one or more mutations which

permit the ADAT to deaminate a target A in DNA. For example, the base editor can comprise all or a portion of an ADAT from *Escherichia coli* (EcTadA) comprising one or more of the following mutations: D108N, A106V, D147Y, E155V, L84F, H123Y, I156F, or a corresponding mutation in another adenosine deaminase.

The adenosine deaminase can be derived from any suitable organism (*e.g.*, *E. coli*). In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli, Staphylococcus aureus, Salmonella typhi, Shewanella putrefaciens, Haemophilus influenzae, Caulobacter crescentus,* or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*. In some embodiments, the adenine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (*e.g.*, mutations in ecTadA). The corresponding residue in any homologous protein can be identified by *e.g.*, sequence alignment and determination of homologous residues. The mutations in any naturally-occurring adenosine deaminase (*e.g.*, having homology to ecTadA) that correspond to any of the mutations described herein (*e.g.*, any of the mutations identified in ecTadA) can be generated accordingly.

Adenosine deaminases

In some embodiments, the fusion proteins as described herein comprise one or more adenosine deaminase domains. In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine. In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine in a deoxyadenosine residue of DNA. The adenosine deaminase may be derived from any suitable organism (*e.g.*, *E. coli*). In some embodiments, the adenine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (*e.g.*, mutations in ecTadA). One of skill in the art will be able to identify the corresponding residue in any homologous protein, *e.g.*, by sequence alignment and determination of homologous residues. Accordingly, one of skill in the art would be able to generate mutations in any naturally-occurring adenosine deaminase (*e.g.*, having homology to ecTadA) that corresponds to any of the mutations described herein, *e.g.*, any of the mutations identified in ecTadA. In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments,

the adenosine deaminase is from *Escherichia coli, Staphylococcus aureus, Salmonella typhi, Shewanella putrefaciens, Haemophilus influenzae, Caulobacter crescentus,* or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*.

Provided and described herein are adenosine deaminase variants that have increased efficiency (>50-60%) and specificity. In particular, the adenosine deaminase variants described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (*i.e.*, "bystanders").

In some embodiments, the adenosine deaminase is a TadA deaminase. In particular embodiments, the TadA is any one of the TadA described in PCT/US2017/045381 (WO 2018/027078), which is incorporated herein by reference in its entirety.

A wild type TadA(wt) adenosine deaminase has the following sequence (also termed TadA reference sequence):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHP
GMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 391)

In some embodiments the adenosine deaminase is a full-length *E. coli* TadA deaminase. For example, in certain embodiments, the adenosine deaminase comprises the amino acid sequence:

MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGA
AGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 392).

In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli (E. coli)*, *Staphylococcus aureus (S. aureus)*, *Salmonella typhimurium* (*S. typhimurium*), *Shewanella putrefaciens* (*S. putrefaciens*), *Haemophilus influenzae* (*H. influenzae*), *Caulobacter crescentus* (*C. crescentus*), *Geobacter sulfurreducens* (*G. sulfurreducens*), or *Bacillus subtilis*. In some embodiments, the adenosine deaminase is from *E. coli*.

It should be appreciated, however, that additional adenosine deaminases useful in the present application would be apparent to the skilled artisan and are within the scope of this disclosure. For example, the adenosine deaminase may be a homolog of adenosine

deaminase acting on tRNA (ADAT). Without limitation, the amino acid sequences of exemplary AD AT homologs include the following:

Staphylococcus aureus (*S. aureus*) TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIA
IERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQS
NFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN (SEQ ID NO: 309)_

Bacillus subtilis (*B. subtilis*) TadA:

MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEA
CKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNH
QAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE (SEQ ID NO: 310)

Salmonella typhimurium (*S. typhimurium*) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGR
HDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGA
AGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV
(SEQ ID NO: 311)

Shewanella putrefaciens (*S. putrefaciens*) TadA:

MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGK
KLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQV
EVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE (SEQ ID NO: 312)

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHA
EIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHF
FDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK (SEQ ID NO: 313)

Caulobacter crescentus (*C. crescentus*) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHA
EIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKF
FAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI (SEQ ID NO: 314)

Geobacter sulfurreducens (*G. sulfurreducens*) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHA EMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDL SADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP (SEQ ID NO: 315)

An embodiment of *E. Coli* TadA (ecTadA) includes the following:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1)

In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth in any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). The disclosure provides any deaminase domains with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 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, or more mutations compared to a reference sequence, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences known in the art or described herein.

It should be appreciated that any of the mutations provided herein (*e.g.*, based on the TadA reference sequence) can be introduced into other adenosine deaminases, such as *E. coli* TadA (ecTadA), *S. aureus* TadA (saTadA), or other adenosine deaminases (*e.g.*, bacterial adenosine deaminases). It would be apparent to the skilled artisan that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein. Thus, any of the mutations identified in the TadA reference sequence can be made in other adenosine deaminases (*e.g.*, ecTada) that have homologous amino acid

residues. It should also be appreciated that any of the mutations provided herein can be made individually or in any combination in the TadA reference sequence or another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises a D108X mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108G, D108N, D108V, D108A, or D108Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase. It should be appreciated, however, that additional deaminases may similarly be aligned to identify homologous amino acid residues that can be mutated as provided herein.

In some embodiments, the adenosine deaminase comprises an A106X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a E155X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a E155D, E155G, or E155V mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises a D147X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D147Y, mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A106X, E155X, or D147X, mutation in the TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA), where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the

adenosine deaminase comprises an E155D, E155G, or E155V mutation. In some embodiments, the adenosine deaminase comprises a D147Y.

It should be appreciated that any of the mutations provided herein (*e.g.*, based on the ecTadA amino acid sequence of TadA reference sequence) may be introduced into other adenosine deaminases, such as *S. aureus* TadA (saTadA), or other adenosine deaminases (*e.g.*, bacterial adenosine deaminases). It would be apparent to the skilled artisan how to are homologous to the mutated residues in ecTadA. Thus, any of the mutations identified in ecTadA may be made in other adenosine deaminases that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein may be made individually or in any combination in ecTadA or another adenosine deaminase.

For example, an adenosine deaminase contains a combination of mutations (*e.g.*, V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; or L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N), and may contain one or more additional mutations. Additional mutations include, for example, a D108N, a A106V, a E155V, and/or a D147Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, an adenosine deaminase comprises the following group of mutations (groups of mutations are separated by a ";") in TadA reference sequence, or corresponding mutations in another adenosine deaminase: D108N and A106V; D108N and E155V; D108N and D147Y; A106V and E155V; A106V and D147Y; E155V and D147Y; D108N, A106V, and E155V; D108N, A106V, and D147Y; D108N, E155V, and D147Y; A106V, E155V, and D147Y; and D108N, A106V, E155V, and D147Y. It should be appreciated, however, that any combination of corresponding mutations provided herein may be made in an adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H8X, T17X, L18X, W23X, L34X, W45X, R51X, A56X, E59X, E85X, M94X, I95X, V102X, F104X, A106X, R107X, D108X, K110X, M118X, N127X, A138X, F149X, M151X, R153X, Q154X, I156X, and/or K157X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y,

T17S, L18E, W23L, L34S, W45L, R51H, A56E, or A56S, E59G, E85K, or E85G, M94L, I95L, V102A, F104L, A106V, R107C, or R107H, or R107P, D108G, or D108N, or D108V, or D108A, or D108Y, K110I, M118K, N127S, A138V, F149Y, M151V, R153C, Q154L, I156D, and/or K157R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a H8X, D108X, and/or N127X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid. In some embodiments, the adenosine deaminase comprises one or more of a H8Y, D108N, and/or N127S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of H8X, R26X, M61X, L68X, M70X, A106X, D108X, A109X, N127X, D147X, R152X, Q154X, E155X, K161X, Q163X, and/or T166X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H8Y, R26W, M61I, L68Q, M70V, A106T, D108N, A109T, N127S, D147Y, R152C, Q154H or Q154R, E155G or E155V or E155D, K161Q, Q163H, and/or T166P mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, D108X, N127X, D147X, R152X, and Q154X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.,* ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, M61X, M70X, D108X, N127X, Q154X, E155X, and Q163X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.,* ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8X, D108X, N127X, E155X, and T166X in TadA reference sequence, or a corresponding mutation or mutations in another

adenosine deaminase (*e.g.*, ecTadA), where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8X, A106X, and D108X, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8X, R26X, L68X, D108X, N127X, D147X, and E155X, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of H8X, R126X, L68X, D108X, N127X, D147X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8X, D108X, A109X, N127X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of H8Y, D108N, N127S, D147Y, R152C, and Q154H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, M61I, M70V, D108N, N127S, Q154R, E155G and Q163H in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, N127S, E155V, and T166P in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the

group consisting of H8Y, A106T, D108N, N127S, E155D, and K161Q in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, seven, or eight mutations selected from the group consisting of H8Y, R26W, L68Q, D108N, N127S, D147Y, and E155V in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA). In some embodiments, the adenosine deaminase comprises one, two, three, four, or five, mutations selected from the group consisting of H8Y, D108N, A109T, N127S, and E155G in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of the or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D108G, or D108V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V and D108N mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises R107C and D108N mutations in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and Q154H mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, N127S, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a H8Y, D108N, and N127S mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A106V, D108N, D147Y, and E155V mutation in TadA reference sequence, or corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of S2X, H8X, I49X, L84X, H123X, N127X, I156X, and/or K160X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence

of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of S2A, H8Y, I49F, L84F, H123Y, N127S, I156F, and/or K160S mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an L84X mutation adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an L84F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an H123X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H123Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an I156X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an I156F mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84X, A106X, D108X, H123X, D147X, E155X, and I156X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2X, I49X, A106X, D108X, D147X, and E155X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group

consisting of H8X, A106X, D108X, N127X, and K160X in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase, where X indicates the presence of any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one, two, three, four, five, six, or seven mutations selected from the group consisting of L84F, A106V, D108N, H123Y, D147Y, E155V, and I156F in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one, two, three, four, five, or six mutations selected from the group consisting of S2A, I49F, A106V, D108N, D147Y, and E155V in TadA reference sequence.

In some embodiments, the adenosine deaminase comprises one, two, three, four, or five mutations selected from the group consisting of H8Y, A106T, D108N, N127S, and K160S in TadA reference sequence, or a corresponding mutation or mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises one or more of a E25X, R26X, R107X, A142X, and/or A143X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of E25M, E25D, E25A, E25R, E25V, E25S, E25Y, R26G, R26N, R26Q, R26C, R26L, R26K, R107P, R107K, R107A, R107N, R107W, R107H, R107S, A142N, A142D, A142G, A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of the mutations described herein corresponding to TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an E25X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an E25M, E25D, E25A, E25R, E25V, E25S, or E25Y mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R26X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises R26G, R26N, R26Q, R26C, R26L, or R26K mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an R107X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R107P, R107K, R107A, R107N, R107W, R107H, or R107S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A142N, A142D, A142G, mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises an A143X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A143D, A143G, A143E, A143L, A143W, A143M, A143S, A143Q, and/or A143R mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase (*e.g.*, ecTadA).

In some embodiments, the adenosine deaminase comprises one or more of a H36X, N37X, P48X, I49X, R51X, M70X, N72X, D77X, E134X, S146X, Q154X, K157X, and/or K161X mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises one or more of H36L, N37T, N37S, P48T, P48L, I49V, R51H, R51L, M70L, N72S, D77G, E134G, S146R, S146C, Q154H, K157N, and/or K161T

mutation in TadA reference sequence, or one or more corresponding mutations in another adenosine deaminase (*e.g.,* ecTadA).

In some embodiments, the adenosine deaminase comprises an H36X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an H36L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an N37X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an N37T or N37S mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an P48T or P48L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R51X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an R51H or R51L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an S146X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an S146R or S146C mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an K157X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a K157N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an P48X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a P48S, P48T, or P48A mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an A142X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a A142N mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an W23X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a W23R or W23L mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In some embodiments, the adenosine deaminase comprises an R152X mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a R152P or R52H mutation in TadA reference sequence, or a corresponding mutation in another adenosine deaminase.

In one embodiment, the adenosine deaminase may comprise the mutations H36L, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, E155V, I156F, and K157N. In some embodiments, the adenosine deaminase comprises the following combination of mutations

relative to TadA reference sequence, where each mutation of a combination is separated by a "_" and each combination of mutations is between parentheses:

(A106V_D108N),

(R107C_D108N),

(H8Y_D108N_N127S_D147Y_Q154H),

(H8Y _D108N_N127S_D147Y_E155V),

(D108N_D147Y_E155V),

(H8Y_D108N_N127S),

(H8Y_D108N_N127S_D147Y_Q154H),

(A106V_D108N_D147Y_E155V),

(D108Q_D147Y_E155V),

(D108M_D147Y_E155V),

(D108L_D147Y_E155V),

(D108K_D147Y_E155V),

(D108I_D147Y_E155V),

(D108F_D147Y_E155V),

(A106V_D108N_D147Y),

(A106V_D108M_D147Y_E155V),

(E59A_A106V_D108N_D147Y_E155V),

(E59A cat dead_A106V_D108N_D147Y_E155V),

(L84F_A106V_D108N_H123Y_D147Y_E155V_I156Y),

(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

(D103A_D104N),

(G22P_D103A_D104N),

(D103A_D104N_S138A),

(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

(E25G_R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

(E25D_R26G_L84F_A106V_R107K_D108N_H123Y_A142N_A143G_D147Y_E155V_I156F),

(R26Q_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

(E25M_R26G_L84F_A106V_R107P_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

(R26C_L84F_A106V_R107H_D108N_H123Y_A142N_D147Y_E155V_I156F),

(L84F_A106V_D108N_H123Y_A142N_A143L_D147Y_E155V_I156F),

(R26G_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),

(E25A_R26G_L84F_A106V_R107N_D108N_H123Y_A142N_A143E_D147Y_E155V_I156F),

(R26G_L84F_A106V_R107H_D108N_H123Y_A142N_A143D_D147Y_E155V_I156F),

(A106V_D108N_A142N_D147Y_E155V),

(R26G_A106V_D108N_A142N_D147Y_E155V),
(E25D_R26G_A106V_R107K_D108N_A142N_A143G_D147Y_E155V),
(R26G_A106V_D108N_R107H_A142N_A143D_D147Y_E155V),
(E25D_R26G_A106V_D108N_A142N_D147Y_E155V),
(A106V_R107K_D108N_A142N_D147Y_E155V),
(A106V_D108N_A142N_A143G_D147Y_E155V),
(A106V_D108N_A142N_A143L_D147Y_E155V),
(H36L_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(N37T_P48T_M70L_L84F_A106V_D108N_H123Y_D147Y_I49V_E155V_I156F),
(N37S_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K161T),
(H36L_L84F_A106V_D108N_H123Y_D147Y_Q154H_E155V_I156F),
(N72S_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F),
(H36L_P48L_L84F_A106V_D108N_H123Y_E134G_D147Y_E155V_I156F),
(H36L_L84F_A106V_D108N_H123Y_ D147Y_E155V_I156F_K157N)
(H36L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F_K161T),
(N37S_R51H_D77G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_K157N),
(D24G_Q71R_L84F_H96L_A106V_D108N_H123Y_D147Y_E155V_I156F_K160E),
(H36L_G67V_L84F_A106V_D108N_H123Y_S146T_D147Y_E155V_I156F),
(Q71L_L84F_A106V_D108N_H123Y_L137M_A143E_D147Y_E155V_I156F),
(E25G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A91T_F104I_A106V_D108N_H123Y_D147Y_E155V_I156F),
(N72D_L84F_A106V_D108N_H123Y_G125A_D147Y_E155V_I156F),
(P48S_L84F_S97C_A106V_D108N_H123Y_D147Y_E155V_I156F),
(W23G_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(D24G_P48L_Q71R_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F_Q159L),
(L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(H36L_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F_K157N),
(N37S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_D147Y_E155V_I156F),
(R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E_K161T),
(L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F_K157N_K160E),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),

(R74A_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(R74Q_L84F_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_R98Q_A106V_D108N_H123Y_D147Y_E155V_I156F),
(L84F_A106V_D108N_H123Y_R129Q_D147Y_E155V_I156F),
(P48S_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F),
(P48S_A142N),
(P48T_I49V_L84F_A106V_D108N_H123Y_A142N_D147Y_E155V_I156F_L157N),
(P48T_I49V_A142N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48S_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48T_I49V_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_ I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_A142N_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F
_K161T),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152H_E155V_I156F
_K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V _I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_E155V_I156F
_K157N),
(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142A_S146C_D147Y_R152P
_E155V_I156F _K157N),

(W23L_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146R_D147Y_E155V_I156F _K161T),
(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V _I156F _K157N),
(H36L_P48A_R51L_L84F_A106V_D108N_H123Y_A142N_S146C_D147Y_R152P_E155 V_I156F _K157N).

In some embodiments, the TadA deaminase is TadA variant. In some embodiments, the TadA variant is TadA*7.10. In particular embodiments, the fusion proteins comprise a single TadA*7.10 domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein comprises TadA*7.10 and TadA(wt), which are capable of forming heterodimers. In one embodiment, a fusion protein as described herein comprises a wild-type TadA linked to TadA*7.10, which is linked to Cas9 nickase.

In some embodiments, TadA*7.10 comprises at least one alteration. In some embodiments, the adenosine deaminase comprises an alteration in the following sequence:

TadA*7.10
MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1)

In some embodiments, TadA*7.10 comprises an alteration at amino acid 82 and/or 166. In particular embodiments, TadA*7.10 comprises one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R. In other embodiments, a variant of TadA*7.10 comprises a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R.

In some embodiments, a variant of TadA*7.10 comprises one or more of alterations selected from the group of L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N. In some embodiments, a variant of TadA*7.10 comprises V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N. In other embodiments, a variant of TadA*7.10 comprises a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G +

Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N.

In some embodiments, an adenosine deaminase variant (*e.g.*, TadA variant) comprises a deletion. In some embodiments, an adenosine deaminase variant comprises a deletion of the C terminus. In particular embodiments, an adenosine deaminase variant comprises a deletion of the C terminus beginning at residue 149, 150, 151, 152, 153, 154, 155, 156, and 157, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, an adenosine deaminase variant (*e.g.*, TadA*8) is a monomer comprising one or more of the following alterations: Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) is a monomer comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) monomer comprising one or more of the following alterations: R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA*8) monomer comprises a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, an adenosine deaminase variant (*e.g.*, MSP828) is a monomer comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D,

F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant (*e.g.*, MSP828) is a monomer comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant (TadA variant) is a monomer comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor of the disclosure comprising an adenosine deaminase variant (*e.g.*, TadA*8) homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*8) each having a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C +

A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, an adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*7.10) each having one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a homodimer comprising two adenosine deaminase variant domains (*e.g.*, MSP828) each having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a homodimer comprising two adenosine deaminase domains (*e.g.*, TadA*7.10) each having a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y;

V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a heterodimer comprising a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, MSP828) having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a wild-type adenosine deaminase domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations Y147T, Y147R, Q154S, Y123H, V82S, T166R, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, a base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising one or more of the following alterations R26C, V88A, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the base editor comprises a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*8) comprising a combination of alterations selected from the group of: R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N; V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N; V88A + T111R + D119N + F149Y; and A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising one or more of the following alterations L36H, I76Y, V82G, Y147T, Y147D, F149Y, Q154S, N157K, and/or D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase variant is a heterodimer comprising a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, MSP828) having the following alterations V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N, relative to TadA*7.10,

the TadA reference sequence, or a corresponding mutation in another TadA. In other embodiments, the adenosine deaminase variant is a heterodimer of a TadA*7.10 domain and an adenosine deaminase variant domain (*e.g.*, TadA*7.10) comprising a combination of alterations selected from the group of: V82G + Y147T + Q154S; I76Y + V82G + Y147T + Q154S; L36H + V82G + Y147T + Q154S + N157K; V82G + Y147D + F149Y + Q154S + D167N; L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N; L36H + I76Y + V82G + Y147T + Q154S + N157K; I76Y + V82G + Y147D + F149Y + Q154S + D167N; L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In some embodiments, the TadA*8 is a variant as shown in **Tables 8A, 10, 11, or 13**. **Tables 8A, 10, 11,** and **13** show certain amino acid position numbers in the TadA amino acid sequence and the amino acids present in those positions in the TadA-7.10 adenosine deaminase. **Tables 8A, 10, 11,** and **13** also show amino acid changes in TadA variants relative to TadA-7.10 following phage-assisted non-continuous evolution (PANCE) and phage-assisted continuous evolution (PACE), as described in M. Richter *et al.*, 2020, *Nature Biotechnology*, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein. In some embodiments, the TadA*8 is TadA*8a, TadA*8b, TadA*8c, TadA*8d, or TadA*8e. In some embodiments, the TadA*8 is TadA*8e.

In particular embodiments, an adenosine deaminase heterodimer can comprise a TadA*8 domain and an adenosine deaminase domain selected from *Staphylococcus aureus (S. aureus) TadA, Bacillus subtilis (B. subtilis) TadA, Salmonella typhimurium (S. typhimurium) TadA, Shewanella putrefaciens (S. putrefaciens) TadA, Haemophilus influenzae F3031 (H. influenzae) TadA, Caulobacter crescentus (C. crescentus) TadA, Geobacter sulfurreducens (G. sulfurreducens)* TadA, or TadA*7.10.

In some embodiments, an adenosine deaminase is a TadA*8. In one embodiment, an adenosine deaminase is a TadA*8 that comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 316)

In some embodiments, the TadA*8 is truncated. In some embodiments, the truncated TadA*8 is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA*8. In some embodiments, the

truncated TadA*8 is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA*8. In some embodiments the adenosine deaminase variant is a full-length TadA*8.

In one embodiment, a fusion protein as described and/or exemplified herein comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (e.g., TadA*8), which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA*8 domain (e.g., provided as a monomer). In other embodiments, the base editor comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

In some embodiments the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24

Table 8A. Additional TadA*8 Variants

		TadA amino acid number									
	TadA	26	88	109	111	119	122	147	149	166	167
	TadA-7.10	R	V	A	T	D	H	Y	F	T	D
PANCE 1					R						
PANCE 2				S/T	R						
PACE	TadA-8a	C		S	R	N	N	D	Y	I	N
	TadA-8b		A	S	R	N	N		Y	I	N
	TadA-8c	C		S	R	N	N		Y	I	N
	TadA-8d		A		R	N			Y		
	TadA-8e			S	R	N	N	D	Y	I	N

In some embodiments, the TadA variant is a variant as shown in **Table 8B**. **Table 8B** shows certain amino acid position numbers in the TadA amino acid sequence and the amino acids present in those positions in the TadA*7.10 adenosine deaminase. In some embodiments, the TadA variant is MSP605, MSP680, MSP823, MSP824, MSP825, MSP827, MSP828, or MSP829. In some embodiments, the TadA variant is MSP828. In some embodiments, the TadA variant is MSP829.

Table 8B. TadA Variants

Variant	TadA Amino Acid Number							
	36	**76**	**82**	**147**	**149**	**154**	**157**	**167**
TadA-7.10	L	I	V	Y	F	Q	N	D
MSP605			G	T		S		
MSP680		Y	G	T		S		
MSP823	H		G	T		S	K	
MSP824			G	D	Y	S		N
MSP825	H		G	D	Y	S	K	N
MSP827	H	Y	G	T		S	K	
MSP828		Y	G	D	Y	S		N
MSP829	H	Y	G	D	Y	S	K	N

In one embodiment, a fusion protein as described herein comprises a wild-type TadA is linked to an adenosine deaminase variant described herein, which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single variant TadA domain (*e.g.*, provided as a monomer). In other embodiments, the fusion protein comprises a variant TadA and TadA(wt), which are capable of forming heterodimers.

In some embodiments, the TadA variant is truncated. In some embodiments, the truncated TadA is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length TadA variant. In some embodiments, the truncated TadA variant is missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length TadA variant. In some embodiments the adenosine deaminase variant is a full-length TadA variant.

In particular embodiments, a TadA*8 comprises one or more mutations at any of the following positions shown in bold. In other embodiments, a TadA*8 comprises one or more mutations at any of the positions shown with underlining:

MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG [50]
LHDPTAHAEI MALRQGGLVM QN**Y**RLIDATL Y**V**TFEPCVMC AGAMIHSRIG [100]
RVVFGVRNAK TGAAGSLMDV LH**Y**PGMNHRV EITEGILADE CAALLC**Y**FFR [150]
MPR**Q**VFNAQK KAQSS**T**D (SEQ ID NO: 1).

For example, the TadA*8 comprises alterations at amino acid position 82 and/or 166 (e.g., V82S, T166R) alone or in combination with any one or more of the following Y147T,

Y147R, Q154S, Y123H, and/or Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA.

In particular embodiments, a combination of alterations is selected from the group of: Y147T + Q154R; Y147T + Q154S; Y147R + Q154S; V82S + Q154S; V82S + Y147R; V82S + Q154R; V82S + Y123H; I76Y + V82S; V82S + Y123H + Y147T; V82S + Y123H + Y147R; V82S + Y123H + Q154R; Y147R + Q154R +Y123H; Y147R + Q154R + I76Y; Y147R + Q154R + T166R; Y123H + Y147R + Q154R + I76Y; V82S + Y123H + Y147R + Q154R; and I76Y + V82S + Y123H + Y147R + Q154R, relative to TadA*7.10, the TadA reference sequence, or a corresponding mutation in another TadA. In some embodiments, an adenosine deaminase comprises one or more of the following alterations: R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, V82T, M94V, P124W, T133K, D139L, D139M, C146R, and A158K.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: V82S + Q154R + Y147R; V82S + Q154R + Y123H; V82S + Q154R + Y147R+ Y123H; Q154R + Y147R + Y123H + I76Y+ V82S; V82S + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; Q154R + Y147R + Y123H + I76Y; V82S + Y147R; V82S + Y147R + Y123H; V82S + Q154R + Y123H; V82S + Q154R + Y147R; V82S + Q154R + Y147R; Q154R + Y147R + Y123H + I76Y; Q154R + Y147R + Y123H + I76Y + V82S; I76Y_V82S_Y123H_Y147R_Q154R; Y147R + Q154R + H123H; and V82S + Q154R.

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: E25F + V82S + Y123H, T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; Q71M + V82S + Y123H + Y147R + Q154R; E25F + V82S + Y123H + T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; V82S + Y123H + P124W + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; R23H + V82S + Y123H + Y147R + Q154R; R21N + V82S + Y123H + Y147R + Q154R; V82S + Y123H + Y147R + Q154R + A158K; N72K + V82S + Y123H + D139L + Y147R +

Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; and M70V + V82S + M94V + Y123H + Y147R + Q154R

In some embodiments, an adenosine deaminase comprises one or more of the following combinations of alterations: Q71M + V82S + Y123H + Y147R + Q154R; E25F + I76Y+ V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D139M + Y147R + Q154R; Y73S + I76Y + V82S + Y123H + Y147R + Q154R; and V82S + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R; N72K_V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; M70V +V82S + M94V + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; and M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R. In some embodiments, the adenosine deaminase is expressed as a monomer. In other embodiments, the adenosine deaminase is expressed as a heterodimer. In some embodiments, the deaminase or other polypeptide sequence lacks a methionine, for example when included as a component of a fusion protein. This can alter the numbering of positions. However, the skilled person will understand that such corresponding mutations refer to the same mutation, e.g., Y73S and Y72S and D139M and D138M.

In some embodiments, the TadA*9 variant is a monomer. In some embodiments, the TadA*9 variant is a heterodimer with a wild-type TadA adenosine deaminase. In some embodiments, the TadA*9 variant is a heterodimer with another TadA variant (e.g., TadA*8, TadA*9). Additional details of TadA*9 adenosine deaminases are described in International PCT Application No. PCT/2020/049975, which is incorporated herein by reference in its entirety. In one embodiment, a fusion protein as described herein comprises a wild-type TadA is linked to an adenosine deaminase variant described herein (*e.g.*, TadA variant),

which is linked to Cas9 nickase. In particular embodiments, the fusion proteins comprise a single TadA variant domain (*e.g.*, provided as a monomer). In other embodiments, the base editor comprises TadA*8 and TadA(wt), which are capable of forming heterodimers.

In particular embodiments, the fusion proteins comprise a single (*e.g.*, provided as a monomer) TadA variant domain. In some embodiments, the TadA variant is linked to a Cas9 nickase. In some embodiments, the fusion proteins described herein comprise as a heterodimer of a wild-type TadA (TadA(wt)) linked to a TadA variant. In other embodiments, the fusion proteins described herein comprise as a heterodimer of a TadA*7.10 linked to a TadA variant. In some embodiments, the fusion protein comprises a TadA variant monomer. In some embodiments, the fusion protein comprises a heterodimer of a TadA variant and a TadA(wt). In some embodiments, the fusion protein comprises a heterodimer of a TadA variant and TadA*7.10. In some embodiments, the fusion protein comprises a heterodimer of two TadA variants. In some embodiments, the TadA variant is selected from **Table 8A, 8B, 9, 10, 11, 12, 13, 14A, 14B, 18, or 20** *infra* or any other TadA variant provided herein.

In some embodiments, the deaminase or other polypeptide sequence lacks a methionine, for example when included as a component of a fusion protein. This can alter the numbering of positions. However, the skilled person will understand that such corresponding mutations refer to the same mutation.

Any of the mutations provided herein and any additional mutations (*e.g.*, based on the ecTadA amino acid sequence) can be introduced into any other adenosine deaminases. Any of the mutations provided herein can be made individually or in any combination in TadA reference sequence or another adenosine deaminase (*e.g.*, ecTadA).

Details of A to G nucleobase editing proteins are described in International PCT Application No. PCT/2017/045381 (WO2018/027078) and Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature, 551, 464-471 (2017), the entire contents of which are hereby incorporated by reference.

High fidelity Cas9 domains

Some aspects of the disclosure provide high fidelity Cas9 domains. In some embodiments, high fidelity Cas9 domains are engineered Cas9 domains comprising one or more mutations that decrease electrostatic interactions between the Cas9 domain and a sugar-

phosphate backbone of a DNA, as compared to a corresponding wild-type Cas9 domain. Without wishing to be bound by any particular theory, high fidelity Cas9 domains that have decreased electrostatic interactions with a sugar-phosphate backbone of DNA may have less off-target effects. In some embodiments, a Cas9 domain (*e.g.*, a wild type Cas9 domain) comprises one or more mutations that decreases the association between the Cas9 domain and a sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and a sugar-phosphate backbone of a DNA by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, or at least 70%.

In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a N497X, a R661X, a Q695X, and/or a Q926X mutation, or a corresponding mutation in any of the amino acid sequences provided herein, wherein X is any amino acid. In some embodiments, any of the Cas9 fusion proteins provided herein comprise one or more of a N497A, a R661A, a Q695A, and/or a Q926A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. In some embodiments, the Cas9 domain comprises a D10A mutation, or a corresponding mutation in any of the amino acid sequences provided herein. Cas9 domains with high fidelity are known in the art and would be apparent to the skilled artisan. For example, Cas9 domains with high fidelity have been described in Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., *et al.* "Rationally engineered Cas9 nucleases with improved specificity." *Science* 351, 84-88 (2015); the entire contents of each are incorporated herein by reference. An Exemplary high fidelity Cas9 domain is provided in the Sequence Listing as SEQ ID NO: 233. In some embodiments, high fidelity Cas9 domains are engineered Cas9 domains comprising one or more mutations that decrease electrostatic interactions between the Cas9 domain and the sugar-phosphate backbone of a DNA, relative to a corresponding wild-type Cas9 domain. High fidelity Cas9 domains that have decreased electrostatic interactions with the sugar-phosphate backbone of DNA have less off-target effects. In some embodiments, the Cas9 domain (e.g., a wild type Cas9 domain (SEQ ID NOs: 197 and 200)) comprises one or more mutations that decrease the association between the Cas9 domain and the sugar-phosphate backbone of a DNA. In some embodiments, a Cas9 domain comprises one or more mutations that decreases the association between the Cas9 domain and the sugar-phosphate backbone of DNA by at least

1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, or at least 70%.

In some embodiments, the modified Cas9 is a high fidelity Cas9 enzyme. In some embodiments, the high fidelity Cas9 enzyme is SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, or hyper accurate Cas9 variant (HypaCas9). The modified Cas9 eSpCas9(1.1) contains alanine substitutions that weaken the interactions between the HNH/RuvC groove and the non-target DNA strand, preventing strand separation and cutting at off-target sites. Similarly, SpCas9-HF1 lowers off-target editing through alanine substitutions that disrupt Cas9's interactions with the DNA phosphate backbone. HypaCas9 contains mutations (SpCas9 N692A/M694A/Q695A/H698A) in the REC3 domain that increase Cas9 proofreading and target discrimination. All three high fidelity enzymes generate less off-target editing than wildtype Cas9.

An exemplary high fidelity Cas9 is provided below. High Fidelity Cas9 domain mutations relative to Cas9 are shown in bold and underlined.

DKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDE
VAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ
LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT
PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTE
ITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT**A**
FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV
TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVL
TLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTILDFL
KSDGFANRNFM**A**LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD
ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ
NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDN
VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVV
GTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSD

KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPI
DFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH
YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD (SEQ ID NO: 233)

Fusion Proteins Comprising a NapDNAbp and a Cytidine Deaminase and/or Adenosine Deaminase

Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain or other nucleic acid programmable DNA binding protein (*e.g.*, Cas12) and one or more cytidine deaminase and/or adenosine deaminase domains. It should be appreciated that the Cas9 domain may be any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein. In some embodiments, any of the Cas9 domains or Cas9 proteins (*e.g.*, dCas9 or nCas9) provided herein may be fused with any of the cytidine deaminases and/or adenosine deaminases provided herein. The domains of the base editors disclosed herein can be arranged in any order.

In some embodiments, the fusion protein comprises the following domains A-C, A-D, or A-E:

NH_2-[A-B-C]-COOH;

NH_2-[A-B-C-D]-COOH; or

NH_2-[A-B-C-D-E]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

an adenosine deaminase domain or an active fragment thereof,

a cytidine deaminase domain or an active fragment thereof, and

wherein B or B and D, each comprises one or more domains having nucleic acid sequence specific binding activity.

In some embodiments, the fusion protein comprises the following structure:

NH_2-[A_n-B_o-C_n]-COOH;

NH_2-[A_n-B_o-C_n-D_o]-COOH; or

NH_2-[A_n-B_o-C_p-D_o-E_q]-COOH;

wherein A and C or A, C, and E, each comprises one or more of the following:

an adenosine deaminase domain or an active fragment thereof,

a cytidine deaminase domain or an active fragment thereof, and

wherein n is an integer: 1, 2, 3, 4, or 5, wherein p is an integer: 0, 1, 2, 3, 4, or 5; wherein q is an integer 0, 1, 2, 3, 4, or 5; and wherein B or B and D each comprises a domain having nucleic acid sequence specific binding activity; and wherein o is an integer: 1, 2, 3, 4, or 5.

For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

NH2-[adenosine deaminase]-[Cas9 domain]-COOH;

NH2-[Cas9 domain]-[adenosine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas9 domain]-COOH;

NH2-[Cas9 domain]-[cytidine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas9 domain]-[adenosine deaminase]-COOH;

NH2-[adenosine deaminase]-[Cas9 domain]-[cytidine deaminase]-COOH;

NH2-[adenosine deaminase]-[cytidine deaminase]-[Cas9 domain]-COOH;

NH2-[cytidine deaminase]-[adenosine deaminase]-[Cas9 domain]-COOH;

NH2-[Cas9 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or

NH2-[Cas9 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, any of the Cas12 domains or Cas12 proteins provided herein may be fused with any of the cytidine or adenosine deaminases provided herein. For example, and without limitation, in some embodiments, the fusion protein comprises the structure:

NH2-[adenosine deaminase]-[Cas12 domain]-COOH;

NH2-[Cas12 domain]-[adenosine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas12 domain]-COOH;

NH2-[Cas12 domain]-[cytidine deaminase]-COOH;

NH2-[cytidine deaminase]-[Cas12 domain]-[adenosine deaminase]-COOH;

NH2-[adenosine deaminase]-[Cas12 domain]-[cytidine deaminase]-COOH;

NH2-[adenosine deaminase]-[cytidine deaminase]-[Cas12 domain]-COOH;

NH2-[cytidine deaminase]-[adenosine deaminase]-[Cas12 domain]-COOH;

NH2-[Cas12 domain]-[adenosine deaminase]-[cytidine deaminase]-COOH; or

NH2-[Cas12 domain]-[cytidine deaminase]-[adenosine deaminase]-COOH.

In some embodiments, the adenosine deaminase is a TadA*8. Exemplary fusion protein structures include the following:

NH2-[TadA*8]-[Cas9 domain]-COOH;

NH2-[Cas9 domain]-[TadA*8]-COOH;

NH2-[TadA*8]-[Cas12 domain]-COOH; or

NH2-[Cas12 domain]-[TadA*8]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein comprises a TadA*8 and a cytidine deaminase and/or an adenosine deaminase. In some embodiments, the TadA*8 is TadA*8.1, TadA*8.2, TadA*8.3, TadA*8.4, TadA*8.5, TadA*8.6, TadA*8.7, TadA*8.8, TadA*8.9, TadA*8.10, TadA*8.11, TadA*8.12, TadA*8.13, TadA*8.14, TadA*8.15, TadA*8.16, TadA*8.17, TadA*8.18, TadA*8.19, TadA*8.20, TadA*8.21, TadA*8.22, TadA*8.23, or TadA*8.24.

Exemplary fusion protein structures include the following:

NH2-[TadA*8]-[Cas9/Cas12]-[adenosine deaminase]-COOH;

NH2-[adenosine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH;

NH2-[TadA*8]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or

NH2-[cytidine deaminase]-[Cas9/Cas12]-[TadA*8]-COOH.

In some embodiments, the adenosine deaminase of the fusion protein comprises a TadA*9 and a cytidine deaminase and/or an adenosine deaminase. Exemplary fusion protein structures include the following:

NH2-[TadA*9]-[Cas9/Cas12]-[adenosine deaminase]-COOH;

NH2-[adenosine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH;

NH2-[TadA*9]-[Cas9/Cas12]-[cytidine deaminase]-COOH; or

NH2-[cytidine deaminase]-[Cas9/Cas12]-[TadA*9]-COOH.

In some embodiments, the fusion protein can comprise a deaminase flanked by an N-terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises a cytidine deaminase flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 or Cas12 polypeptide. In some embodiments, the fusion protein comprises an adenosine deaminase flanked by an N- terminal fragment and a C-terminal fragment of a Cas9 or Cas 12 polypeptide.

In some embodiments, the fusion proteins comprising a cytidine deaminase or adenosine deaminase and a napDNAbp (*e.g.*, Cas9 or Cas12 domain) do not include a linker sequence. In some embodiments, a linker is present between the cytidine or adenosine deaminase and the napDNAbp. In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker. In some embodiments, cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For

example, in some embodiments the cytidine or adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

It should be appreciated that the fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein may comprise inhibitors, cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags , biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

Exemplary, yet nonlimiting, fusion proteins are described in International PCT Application Nos. PCT/2017/044935, PCT/US2019/044935, and PCT/US2020/016288, each of which is incorporated herein by reference for its entirety.

Fusion proteins comprising a nuclear localization sequence (NLS)

In some embodiments, the fusion proteins provided herein further comprise one or more (*e.g.*, 2, 3, 4, 5) nuclear targeting sequences, for example a nuclear localization sequence (NLS). In one embodiment, a bipartite NLS is used. In some embodiments, a NLS comprises an amino acid sequence that facilitates the importation of a protein, that comprises an NLS, into the cell nucleus (*e.g.*, by nuclear transport). In some embodiments, any of the fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the C-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the C-terminus of an nCas9 domain or a dCas9 domain. In some embodiments, the NLS is fused to the N-terminus of the adenosine deaminase. In some embodiments, the NLS is fused to the C-terminus of the adenosine deaminase. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some

embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. Additional nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises the amino acid sequence PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 328), KRTADGSEFESPKKKRKV (SEQ ID NO: 190), KRPAATKKAGQAKKKK (SEQ ID NO: 191), KKTELQTTNAENKTKKL (SEQ ID NO: 192), KRGINDRNFWRGENGRKTR (SEQ ID NO: 193), RKSGKIAAIVVKRPRKPKKKRKV (SEQ ID NO: 329), or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 196).

In some embodiments, the fusion proteins comprising an adenosine deaminase, a Cas9 domain, and an NLS do not comprise a linker sequence. In some embodiments, linker sequences between one or more of the domains or proteins (*e.g.*, adenosine deaminase, Cas9 domain or NLS) are present. In some embodiments, a linker is present between the adenosine deaminase domains and the napDNAbp. In some embodiments, the "-" used in the general architecture below indicates the presence of an optional linker. In some embodiments, the adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein. For example, in some embodiments the adenosine deaminase and the napDNAbp are fused via any of the linkers provided herein.

In some embodiments, the general architecture of exemplary napDNAbp (*e.g.*, Cas9) fusion proteins with an adenosine deaminase and a napDNAbp (*e.g.*, Cas9) domain comprises any one of the following structures, where NLS is a nuclear localization sequence (*e.g.*, any NLS provided herein), NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein:

NH_2-NLS-[adenosine deaminase]-[napDNAbp domain]-COOH;

NH_2-NLS [napDNAbp domain]-[adenosine deaminase]-COOH;

NH_2-[adenosine deaminase]-[napDNAbp domain]-NLS-COOH;

NH_2-[napDNAbp domain]-[adenosine deaminase]-NLS-COOH;

In some embodiments, the NLS is present in a linker or the NLS is flanked by linkers, for example described herein. A bipartite NLS comprises two basic amino acid clusters, which are separated by a relatively short spacer sequence (hence bipartite - 2 parts, while monopartite NLSs are not). The NLS of nucleoplasmin, KR[PAATKKAGQA]KKKK (SEQ ID

NO: 191), is the prototype of the ubiquitous bipartite signal: two clusters of basic amino acids, separated by a spacer of about 10 amino acids. The sequence of an exemplary bipartite NLS follows: PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 328).

A vector that encodes a CRISPR enzyme comprising one or more nuclear localization sequences (NLSs) can be used. For example, there can be or be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs used. A CRISPR enzyme can comprise the NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 NLSs at or near the carboxy-terminus, or any combination thereof (*e.g.*, one or more NLS at the amino-terminus and one or more NLS at the carboxy terminus). When more than one NLS is present, each can be selected independently of others, such that a single NLS can be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies.

CRISPR enzymes used in the methods can comprise about 6 NLSs. An NLS is considered near the N- or C-terminus when the nearest amino acid to the NLS is within about 50 amino acids along a polypeptide chain from the N- or C-terminus, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, or 50 amino acids.

BASE EDITOR SYSTEM

Provided herein are systems, compositions, and methods for editing a nucleobase using a base editor system. In some embodiments, the base editor system comprises (1) a base editor (BE) comprising a polynucleotide programmable nucleotide binding domain and a nucleobase editing domain (*e.g.*, a deaminase domain) for editing the nucleobase; and (2) a guide polynucleotide (*e.g.*, guide RNA) in conjunction with the polynucleotide programmable nucleotide binding domain. In some embodiments, the base editor system is an adenosine base editor (ABE). In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable DNA binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is a polynucleotide programmable RNA binding domain. In some embodiments, the nucleobase editing domain is a deaminase domain. In some embodiments, a deaminase domain can be an adenine deaminase or an adenosine deaminase. In some embodiments, the adenosine base editor can deaminate adenine in DNA.

In some embodiments, a base editing system as provided herein provides a new approach to genome editing that uses a fusion protein containing a catalytically defective *Streptococcus pyogenes* Cas9, a deaminase (*e.g.*, adenosine deaminase), and an inhibitor of

base excision repair to induce programmable, single nucleotide (C→T or A→G) changes in DNA without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

Details of nucleobase editing proteins are described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

Use of the base editor system provided herein comprises the steps of: (a) contacting a target nucleotide sequence of a polynucleotide (*e.g.*, double- or single stranded DNA or RNA) of a subject with a base editor system comprising a nucleobase editor (*e.g.*, an adenosine base editor) and a guide polynucleic acid (*e.g.*, gRNA), wherein the target nucleotide sequence comprises a targeted nucleobase pair; (b) inducing strand separation of said target region; (c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase; and (d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase. It should be appreciated that in some embodiments, step (b) is omitted. In some embodiments, said targeted nucleobase pair is a plurality of nucleobase pairs in one or more genes. In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is located in the same gene. In some embodiments, the plurality of nucleobase pairs is located in one or more genes, wherein at least one gene is located in a different locus.

In some embodiments, the cut single strand (nicked strand) is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the base editor comprises a Cas9

domain. In some embodiments, the first base is adenine, and the second base is not a G, C, A, or T. In some embodiments, the second base is inosine.

In some embodiments, a single guide polynucleotide may be utilized to target a deaminase to a target nucleic acid sequence. In some embodiments, a single pair of guide polynucleotides may be utilized to target different deaminases to a target nucleic acid sequence.

The nucleobase components and the polynucleotide programmable nucleotide binding component of a base editor system may be associated with each other covalently or non-covalently. For example, in some embodiments, the deaminase domain can be targeted to a target nucleotide sequence by a polynucleotide programmable nucleotide binding domain. In some embodiments, the polynucleotide programmable nucleotide binding domain is non-covalently associated with or attached to the deaminase domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain. In some embodiments, a polynucleotide programmable nucleotide binding domain can target a deaminase domain to a target nucleotide sequence by non-covalently interacting with or associating with the deaminase domain. For example, in some embodiments, the nucleobase editing component, *e.g.*, the deaminase component can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a steril alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

A base editor system may further comprise a guide polynucleotide component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. In some embodiments, a deaminase domain can be targeted to a target nucleotide sequence by a guide polynucleotide. For example, in some embodiments, the nucleobase editing component of the base editor system, *e.g.*, the deaminase component, can comprise an additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (*e.g.*, a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the deaminase domain. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polypeptide. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

In some embodiments, a base editor system can further comprise an inhibitor of base excision repair (BER) component. It should be appreciated that components of the base editor system may be associated with each other via covalent bonds, noncovalent interactions, or any combination of associations and interactions thereof. The inhibitor of BER component may comprise a base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be a uracil DNA glycosylase inhibitor (UGI). In some embodiments, the inhibitor of base excision repair can be an inosine base excision repair inhibitor. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the polynucleotide programmable nucleotide binding domain. In some

embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can be fused or linked to a deaminase domain and an inhibitor of base excision repair. In some embodiments, a polynucleotide programmable nucleotide binding domain can target an inhibitor of base excision repair to a target nucleotide sequence by non-covalently interacting with or associating with the inhibitor of base excision repair. For example, in some embodiments, the inhibitor of base excision repair component can comprise an additional heterologous portion or domain that is capable of interacting with, associating with, or capable of forming a complex with an additional heterologous portion or domain that is part of a polynucleotide programmable nucleotide binding domain. In some embodiments, the inhibitor of base excision repair can be targeted to the target nucleotide sequence by the guide polynucleotide. For example, in some embodiments, the inhibitor of base excision repair can comprise an additional heterologous portion or domain (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) that is capable of interacting with, associating with, or capable of forming a complex with a portion or segment (*e.g.*, a polynucleotide motif) of a guide polynucleotide. In some embodiments, the additional heterologous portion or domain of the guide polynucleotide (*e.g.*, polynucleotide binding domain such as an RNA or DNA binding protein) can be fused or linked to the inhibitor of base excision repair. In some embodiments, the additional heterologous portion may be capable of binding to, interacting with, associating with, or forming a complex with a polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a guide polynucleotide. In some embodiments, the additional heterologous portion may be capable of binding to a polypeptide linker. In some embodiments, the additional heterologous portion may be capable of binding to a polynucleotide linker. The additional heterologous portion may be a protein domain. In some embodiments, the additional heterologous portion may be a K Homology (KH) domain, a MS2 coat protein domain, a PP7 coat protein domain, a SfMu Com coat protein domain, a sterile alpha motif, a telomerase Ku binding motif and Ku protein, a telomerase Sm7 binding motif and Sm7 protein, or an RNA recognition motif.

In some embodiments, the base editor inhibits base excision repair (BER) of the edited strand. In some embodiments, the base editor protects or binds the non-edited strand. In some embodiments, the base editor comprises UGI activity. In some embodiments, the base editor comprises a catalytically inactive inosine-specific nuclease. In some embodiments, the

base editor comprises nickase activity. In some embodiments, the intended edit of base pair is upstream of a PAM site. In some embodiments, the intended edit of base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edit of base-pair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site.

In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker or a spacer. In some embodiments, the linker or spacer is 1-25 amino acids in length. In some embodiments, the linker or spacer is 5-20 amino acids in length. In some embodiments, the linker or spacer is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (*e.g.*, a "deamination window"). In some embodiments, a target can be within a 4 base region. In some embodiments, such a defined target region can be approximately 15 bases upstream of the PAM. See Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1- 10 nucleotides. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edit of base pair is within the target window. In some embodiments, the target window comprises the intended edit of base pair. In some embodiments, the method is performed using any of the base editors provided herein. In some embodiments, a target window is a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base

regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence. For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a polynucleotide programmable nucleotide binding domain. In some embodiments, an NLS of the base editor is localized C-terminal to a polynucleotide programmable nucleotide binding domain.

Other exemplary features that can be present in a base editor as disclosed herein are localization sequences, such as cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

In some embodiments, the adenosine base editor (ABE) can deaminate adenine in DNA. In some embodiments, ABE is generated by replacing APOBEC1 component of BE3 with natural or engineered E. coli TadA, human ADAR2, mouse ADA, or human ADAT2. In some embodiments, ABE comprises evolved TadA variant. In some embodiments, the ABE is ABE 1.2 (TadA*-XTEN-nCas9-NLS). In some embodiments, TadA* comprises A106V and D108N mutations.

In some embodiments, the ABE is a second-generation ABE. In some embodiments, the ABE is ABE2.1, which comprises additional mutations D147Y and E155V in TadA* (TadA*2.1). In some embodiments, the ABE is ABE2.2, ABE2.1 fused to catalytically inactivated version of human alkyl adenine DNA glycosylase (AAG with E125Q mutation). In some embodiments, the ABE is ABE2.3, ABE2.1 fused to catalytically inactivated version of E. coli Endo V (inactivated with D35A mutation). In some embodiments, the ABE is ABE2.6 which has a linker twice as long (32 amino acids, (SGGS)2 (SEQ ID NO: 330)-

XTEN-(SGGS)2 (SEQ ID NO: 330)) as the linker in ABE2.1. In some embodiments, the ABE is ABE2.7, which is ABE2.1 tethered with an additional wild-type TadA monomer. In some embodiments, the ABE is ABE2.8, which is ABE2.1 tethered with an additional TadA*2.1 monomer. In some embodiments, the ABE is ABE2.9, which is a direct fusion of evolved TadA (TadA*2.1) to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.10, which is a direct fusion of wild-type TadA to the N-terminus of ABE2.1. In some embodiments, the ABE is ABE2.11, which is ABE2.9 with an inactivating E59A mutation at the N-terminus of TadA* monomer. In some embodiments, the ABE is ABE2.12, which is ABE2.9 with an inactivating E59A mutation in the internal TadA* monomer.

In some embodiments, the ABE is a fifth generation ABE. In some embodiments, the ABE is ABE5.1, which is generated by importing a consensus set of mutations from surviving clones (H36L, R51L, S146C, and K157N) into ABE3.1. In some embodiments, the ABE is ABE5.3, which has a heterodimeric construct containing wild-type E. coli TadA fused to an internal evolved TadA*. In some embodiments, the ABE is ABE5.2, ABE5.4, ABE5.5, ABE5.6, ABE5.7, ABE5.8, ABE5.9, ABE5.10, ABE5.11, ABE5.12, ABE5.13, or ABE5.14, as shown in **Table 9** below. In some embodiments, the ABE is a sixth generation ABE. In some embodiments, the ABE is ABE6.1, ABE6.2, ABE6.3, ABE6.4, ABE6.5, or ABE6.6, as shown in **Table 9** below. In some embodiments, the ABE is a seventh generation ABE. In some embodiments, the ABE is ABE7.1, ABE7.2, ABE7.3, ABE7.4, ABE7.5, ABE7.6, ABE7.7, ABE7.8, ABE 7.9, or ABE7.10, as shown in **Table 9** below.

In some embodiments, the adenosine base editor (ABE) can deaminate adenine in DNA. In some embodiments, the ABE is a an ABE as shown in **Table 9** below.

Table 9. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
ABE1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
ABE1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
ABE2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
ABE3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
ABE4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
ABE5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
ABE5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
ABE6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
ABE6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K

	23	26	36	37	48	49	51	72	84	87	106	108	123	125	142	146	147	152	155	156	157	161
ABE7.3	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
ABE7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
ABE7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
ABE7.8	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

In some embodiments, the base editor is an eighth generation ABE (ABE8). In some embodiments, the ABE8 contains a TadA*8 variant. In some embodiments, the ABE8 has a monomeric construct containing a TadA*8 variant ("ABE8.x-m"). In some embodiments, the ABE8 is ABE8.1-m, which has a monomeric construct containing TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-m, which has a monomeric construct containing TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-m, which has a monomeric construct containing TadA*7.10 with a Q154S mutation (TadA*8.3).

In some embodiments, the ABE8 is ABE8.4-m, which has a monomeric construct containing TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-m, which has a monomeric construct containing TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-m, which has a monomeric construct containing TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-m, which has a monomeric construct containing TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-m, which has a monomeric construct containing TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-m, which has a monomeric construct containing TadA*7.10 with Y147T and Q154S mutations (TadA*8.12).

In some embodiments, the ABE8 is ABE8.13-m, which has a monomeric construct containing TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-m, which has a monomeric construct containing TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-m, which has a monomeric construct containing TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-m, which has a monomeric construct containing TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-m, which has a monomeric construct containing TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-m, which has a monomeric construct containing TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-m, which has a monomeric construct containing TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-m, which has a monomeric construct containing TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing wild-type E. coli TadA fused to a TadA*8 variant ("ABE8.x-d"). In some embodiments, the ABE8 is ABE8.1-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is

ABE8.4-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-d, which has heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-d, which has a heterodimeric construct containing wild-

type E. coli TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-d, which has a heterodimeric construct containing wild-type E. coli TadA fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE8 has a heterodimeric construct containing TadA*7.10 fused to a TadA*8 variant ("ABE8.x-7"). In some embodiments, the ABE8 is ABE8.1-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147T mutation (TadA*8.1). In some embodiments, the ABE8 is ABE8.2-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y147R mutation (TadA*8.2). In some embodiments, the ABE8 is ABE8.3-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154S mutation (TadA*8.3). In some embodiments, the ABE8 is ABE8.4-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Y123H mutation (TadA*8.4). In some embodiments, the ABE8 is ABE8.5-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a V82S mutation (TadA*8.5). In some embodiments, the ABE8 is ABE8.6-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a T166R mutation (TadA*8.6). In some embodiments, the ABE8 is ABE8.7-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with a Q154R mutation (TadA*8.7). In some embodiments, the ABE8 is ABE8.8-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and Y123H mutations (TadA*8.8). In some embodiments, the ABE8 is ABE8.9-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R and I76Y mutations (TadA*8.9). In some embodiments, the ABE8 is ABE8.10-7, which has a

heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R, Q154R, and T166R mutations (TadA*8.10). In some embodiments, the ABE8 is ABE8.11-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154R mutations (TadA*8.11). In some embodiments, the ABE8 is ABE8.12-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147T and Q154S mutations (TadA*8.12). In some embodiments, the ABE8 is ABE8.13-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y123H (Y123H reverted from H123Y), Y147R, Q154R and I76Y mutations (TadA*8.13). In some embodiments, the ABE8 is ABE8.14-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y and V82S mutations (TadA*8.14). In some embodiments, the ABE8 is ABE8.15-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y147R mutations (TadA*8.15). In some embodiments, the ABE8 is ABE8.16-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Y147R mutations (TadA*8.16). In some embodiments, the ABE8 is ABE8.17-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154R mutations (TadA*8.17). In some embodiments, the ABE8 is ABE8.18-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y) and Q154R mutations (TadA*8.18). In some embodiments, the ABE8 is ABE8.19-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.19). In some embodiments, the ABE8 is ABE8.20-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with I76Y, V82S, Y123H (Y123H reverted from H123Y), Y147R and Q154R mutations (TadA*8.20). In some embodiments, the ABE8 is ABE8.21-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with Y147R and Q154S mutations (TadA*8.21). In some embodiments, the ABE8 is ABE8.22-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Q154S mutations (TadA*8.22). In some embodiments, the ABE8 is ABE8.23-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S and Y123H (Y123H reverted from H123Y) mutations (TadA*8.23). In some embodiments, the ABE8 is ABE8.24-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V82S, Y123H (Y123H reverted from H123Y), and Y147T mutations (TadA*8.24).

In some embodiments, the ABE is ABE8.1-m, ABE8.2-m, ABE8.3-m, ABE8.4-m, ABE8.5-m, ABE8.6-m, ABE8.7-m, ABE8.8-m, ABE8.9-m, ABE8.10-m, ABE8.11-m, ABE8.12-m, ABE8.13-m, ABE8.14-m, ABE8.15-m, ABE8.16-m, ABE8.17-m, ABE8.18-m, ABE8.19-m, ABE8.20-m, ABE8.21-m, ABE8.22-m, ABE8.23-m, ABE8.24-m, ABE8.1-d, ABE8.2-d, ABE8.3-d, ABE8.4-d, ABE8.5-d, ABE8.6-d, ABE8.7-d, ABE8.8-d, ABE8.9-d, ABE8.10-d, ABE8.11-d, ABE8.12-d, ABE8.13-d, ABE8.14-d, ABE8.15-d, ABE8.16-d, ABE8.17-d, ABE8.18-d, ABE8.19-d, ABE8.20-d, ABE8.21-d, ABE8.22-d, ABE8.23-d, or ABE8.24-d as shown in **Table 10** below.

Table 10: Adenosine Deaminase Base Editor 8 (ABE8) Variants

ABE8	Adenosine Deaminase	Adenosine Deaminase Description
ABE8.1-m	TadA*8.1	Monomer_TadA*7.10 + Y147T
ABE8.2-m	TadA*8.2	Monomer_TadA*7.10 + Y147R
ABE8.3-m	TadA*8.3	Monomer_TadA*7.10 + Q154S
ABE8.4-m	TadA*8.4	Monomer_TadA*7.10 + Y123H
ABE8.5-m	TadA*8.5	Monomer_TadA*7.10 + V82S
ABE8.6-m	TadA*8.6	Monomer_TadA*7.10 + T166R
ABE8.7-m	TadA*8.7	Monomer_TadA*7.10 + Q154R
ABE8.8-m	TadA*8.8	Monomer_TadA*7.10 + Y147R_Q154R_Y123H
ABE8.9-m	TadA*8.9	Monomer_TadA*7.10 + Y147R_Q154R_I76Y
ABE8.10-m	TadA*8.10	Monomer_TadA*7.10 + Y147R_Q154R_T166R
ABE8.11-m	TadA*8.11	Monomer_TadA*7.10 + Y147T_Q154R
ABE8.12-m	TadA*8.12	Monomer_TadA*7.10 + Y147T_Q154S
ABE8.13-m	TadA*8.13	Monomer_TadA*7.10 + Y123H_Y147R_Q154R_I76Y
ABE8.14-m	TadA*8.14	Monomer_TadA*7.10 + I76Y_V82S
ABE8.15-m	TadA*8.15	Monomer_TadA*7.10 + V82S_Y147R
ABE8.16-m	TadA*8.16	Monomer_TadA*7.10 + V82S_Y123H_Y147R
ABE8.17-m	TadA*8.17	Monomer_TadA*7.10 + V82S_Q154R
ABE8.18-m	TadA*8.18	Monomer_TadA*7.10 + V82S_Y123H_Q154R
ABE8.19-m	TadA*8.19	Monomer_TadA*7.10 + V82S_Y123H_Y147R_Q154R
ABE8.20-m	TadA*8.20	Monomer_TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R

ABE8.21-m	TadA*8.21	Monomer_TadA*7.10 + Y147R_Q154S
ABE8.22-m	TadA*8.22	Monomer_TadA*7.10 + V82S_Q154S
ABE8.23-m	TadA*8.23	Monomer_TadA*7.10 + V82S_Y123H
ABE8.24-m	TadA*8.24	Monomer_TadA*7.10 + V82S_Y123H_Y147T
ABE8.1-d	TadA*8.1	Heterodimer_(WT) + (TadA*7.10 + Y147T)
ABE8.2-d	TadA*8.2	Heterodimer_(WT) + (TadA*7.10 + Y147R)
ABE8.3-d	TadA*8.3	Heterodimer_(WT) + (TadA*7.10 + Q154S)
ABE8.4-d	TadA*8.4	Heterodimer_(WT) + (TadA*7.10 + Y123H)
ABE8.5-d	TadA*8.5	Heterodimer_(WT) + (TadA*7.10 + V82S)
ABE8.6-d	TadA*8.6	Heterodimer_(WT) + (TadA*7.10 + T166R)
ABE8.7-d	TadA*8.7	Heterodimer_(WT) + (TadA*7.10 + Q154R)
ABE8.8-d	TadA*8.8	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
ABE8.9-d	TadA*8.9	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
ABE8.10-d	TadA*8.10	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
ABE8.11-d	TadA*8.11	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154R)
ABE8.12-d	TadA*8.12	Heterodimer_(WT) + (TadA*7.10 + Y147T_Q154S)
ABE8.13-d	TadA*8.13	Heterodimer_(WT) + (TadA*7.10 + Y123H_Y147T_Q154R_I76Y)
ABE8.14-d	TadA*8.14	Heterodimer_(WT) + (TadA*7.10 + I76Y_V82S)
ABE8.15-d	TadA*8.15	Heterodimer_(WT) + (TadA*7.10 + V82S_ Y147R)
ABE8.16-d	TadA*8.16	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147R)
ABE8.17-d	TadA*8.17	Heterodimer_(WT) + (TadA*7.10 + V82S_Q154R)
ABE8.18-d	TadA*8.18	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Q154R)
ABE8.19-d	TadA*8.19	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147R_Q154R)
ABE8.20-d	TadA*8.20	Heterodimer_(WT) + (TadA*7.10 + I76Y_V82S_Y123H_Y147R_Q154R)
ABE8.21-d	TadA*8.21	Heterodimer_(WT) + (TadA*7.10 + Y147R_Q154S)
ABE8.22-d	TadA*8.22	Heterodimer_(WT) + (TadA*7.10 + V82S_Q154S)
ABE8.23-d	TadA*8.23	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H)
ABE8.24-d	TadA*8.24	Heterodimer_(WT) + (TadA*7.10 + V82S_Y123H_Y147T)

In some embodiments, the ABE8 is ABE8a-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-m, which has a monomeric construct containing TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-m, which has a monomeric construct containing TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some

embodiments, the ABE8 is ABE8d-m, which has a monomeric construct containing TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-m, which has a monomeric construct containing TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-d, which has a heterodimeric construct containing wild-type *E. coli* TadA fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE8 is ABE8a-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8a). In some embodiments, the ABE8 is ABE8b-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V88A, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8b). In some embodiments, the ABE8 is ABE8c-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with R26C, A109S, T111R, D119N, H122N, F149Y, T166I, and D167N mutations (TadA*8c). In some embodiments, the ABE8 is ABE8d-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with V88A, T111R, D119N, and F149Y mutations (TadA*8d). In some embodiments, the ABE8 is ABE8e-7, which has a heterodimeric construct containing TadA*7.10 fused to TadA*7.10 with A109S, T111R, D119N, H122N, Y147D, F149Y, T166I, and D167N mutations (TadA*8e).

In some embodiments, the ABE is ABE8a-m, ABE8b-m, ABE8c-m, ABE8d-m, ABE8e-m, ABE8a-d, ABE8b-d, ABE8c-d, ABE8d-d, or ABE8e-d, as shown in **Table 11**

below. In some embodiments, the ABE is ABE8e-m or ABE8e-d. ABE8e shows efficient adenine base editing activity and low indel formation when used with Cas homologues other than SpCas9, for example, SaCas9, SaCas9-KKH, Cas12a homologues, *e.g.*, LbCas12a, enAs-Cas12a, SpCas9-NG and circularly permuted CP1028-SpCas9 and CP1041-SpCas9. In addition to the mutations shown for ABE8e in **Table 11**, off-target RNA and DNA editing were reduced by introducing a V106W substitution into the TadA domain (as described in M. Richter *et al.*, 2020, Nature Biotechnology, doi.org/10.1038/s41587-020-0453-z, the entire contents of which are incorporated by reference herein).

Table 11: Additional Adenosine Deaminase Base Editor 8 Variants

ABE8 Base Editor	Adenosine Deaminase	Adenosine Deaminase Description
ABE8a-m	TadA*8a	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8b-m	TadA*8b	Monomer_TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8c-m	TadA*8c	Monomer_TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N
ABE8d-m	TadA*8d	Monomer_TadA*7.10 + V88A + T111R + D119N + F149Y
ABE8e-m	TadA*8e	Monomer_TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N
ABE8a-d	TadA*8a	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)
ABE8b-d	TadA*8b	Heterodimer_(WT) + (TadA*7.10 + V88A + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8c-d	TadA*8c	Heterodimer_(WT) + (TadA*7.10 + R26C + A109S + T111R + D119N + H122N + F149Y + T166I + D167N)
ABE8d-d	TadA*8d	Heterodimer_(WT) + (TadA*7.10 + V88A + T111R + D119N + F149Y)
ABE8e-d	TadA*8e	Heterodimer_(WT) + (TadA*7.10 + A109S + T111R + D119N + H122N + Y147D + F149Y + T166I + D167N)

In some embodiments, base editors (*e.g.*, ABE8) are generated by cloning an adenosine deaminase variant (*e.g.*, TadA*8) into a scaffold that includes a circular permutant Cas9 (*e.g.*, CP5 or CP6) and a bipartite nuclear localization sequence. In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP5 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP5 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.*, ABE7.9, ABE7.10, or ABE8) is an NGC PAM CP6 variant (*S. pyogenes* Cas9 or spVRQR Cas9). In some embodiments, the base editor (*e.g.* ABE7.9, ABE7.10, or ABE8) is an AGA PAM CP6 variant (*S. pyogenes* Cas9 or spVRQR Cas9).

In some embodiments, the ABE has a genotype as shown in **Table 12** below.

Table 12. Genotypes of ABEs

	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	161
ABE7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
ABE7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

As shown in **Table 13** below, genotypes of 40 ABE8s are described. Residue positions in the evolved *E. coli* TadA portion of ABE are indicated. Mutational changes in ABE8 are shown when distinct from ABE7.10 mutations. In some embodiments, the ABE has a genotype of one of the ABEs as shown in **Table 13** below.

Table 13. Residue Identity in Evolved TadA

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	**R**	**L**	**A**	**L**	**I**	**V**	**F**	**V**	**N**	**Y**	**C**	**Y**	**P**	**Q**	**V**	**F**	**N**	**T**
ABE8.1-m												T						
ABE8.2-m												R						
ABE8.3-m														S				
ABE8.4-m										H								
ABE8.5-m						S												
ABE8.6-m																		R
ABE8.7-m														R				
ABE8.8-m										H		R		R				
ABE8.9-m					Y							R		R				
ABE8.10-m												R		R				R
ABE8.11-m												T		R				
ABE8.12-m												T		S				
ABE8.13-m					Y					H		R		R				
ABE8.14-m					Y	S												
ABE8.15-m						S						R						
ABE8.16-m						S				H		R						
ABE8.17-m						S								R				
ABE8.18-m						S				H				R				
ABE8.19-m						S				H		R		R				
ABE8.20-m					Y	S				H		R		R				
ABE8.21-m												R		S				
ABE8.22-m						S								S				
ABE8.23-m						S				H								
ABE8.24-m						S				H		T						
ABE8.1-d												T						

	23	36	48	51	76	82	84	106	108	123	146	147	152	154	155	156	157	166
ABE7.10	**R**	**L**	**A**	**L**	**I**	**V**	**F**	**V**	**N**	**Y**	**C**	**Y**	**P**	**Q**	**V**	**F**	**N**	**T**
ABE8.2-d												R						
ABE8.3-d														S				
ABE8.4-d										H								
ABE8.5-d						S												
ABE8.6-d																		R
ABE8.7-d														R				
ABE8.8-d										H		R		R				
ABE8.9-d					Y							R		R				
ABE8.10-d												R		R				R
ABE8.11-d												T		R				
ABE8.12-d												T		S				
ABE8.13-d					Y					H		R		R				
ABE8.14-d					Y	S												
ABE8.15-d						S						R						
ABE8.16-d						S				H		R						
ABE8.17-d						S								R				
ABE8.18-d						S				H				R				
ABE8.19-d						S				H		R		R				
ABE8.20-d					Y	S				H		R		R				
ABE8.21-d												R		S				
ABE8.22-d						S								S				
ABE8.23-d						S				H								
ABE8.24-d						S				H		T						

In some embodiments, the base editor is ABE8.1, which comprises or consists essentially of the following sequence or a fragment thereof having adenosine deaminase activity:

ABE8.1_Y147T_CP5_NGC PAM_monomer

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCTFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSES***
ATPESSGGSSGGSEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFMQPT
VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK
YSLFELENGRKRMLASAKFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF
KYFDTTIARKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGSGGSGGSGGSGGSGGS*

*GGM*DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE
ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL
GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED
IVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENT
QLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGK
SDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQIT
KHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLN
AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEGADKRTADGSEFESPKKKRKV

(SEQ ID NO: 331)

In the above sequence, the plain text denotes an adenosine deaminase sequence, bold sequence indicates sequence derived from Cas9, the italicized sequence denotes a linker sequence, and the underlined sequence denotes a bipartite nuclear localization sequence. Other ABE8 sequences are provided in the attached sequence listing (SEQ ID NOs: 332-354).

In some embodiments, the base editor includes an adenosine deaminase variant comprising an amino acid sequence, which contains alterations relative to an ABE 7*10 reference sequence, as described herein. The term "monomer" as used in **Table 14A** refers to a monomeric form of TadA*7.10 comprising the alterations described. The term "heterodimer" as used in **Table 14A** refers to the specified wild-type *E. coli* TadA adenosine deaminase fused to a TadA*7.10 comprising the alterations as described.

Table 14A. Adenosine Deaminase Base Editor Variants

ABE	Adenosine Deaminase	Adenosine Deaminase Description
ABE-605m	MSP605	monomer_TadA*7.10 + V82G + Y147T + Q154S
ABE-680m	MSP680	monomer_TadA*7.10 + I76Y + V82G + Y147T + Q154S

ABE-823m	MSP823	monomer_TadA*7.10 + L36H + V82G + Y147T + Q154S + N157K
ABE-824m	MSP824	monomer_TadA*7.10 + V82G + Y147D + F149Y + Q154S + D167N
ABE-825m	MSP825	monomer_TadA*7.10 + L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N
ABE-827m	MSP827	monomer_TadA*7.10 + L36H + I76Y + V82G + Y147T + Q154S + N157K
ABE-828m	MSP828	monomer_TadA*7.10 + I76Y + V82G + Y147D + F149Y + Q154S + D167N
ABE-829m	MSP829	monomer_TadA*7.10 + L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N
ABE-605d	MSP605	heterodimer_(WT)+(TadA*7.10 + V82G + Y147T + Q154S)
ABE-680d	MSP680	heterodimer_(WT)+(TadA*7.10 + I76Y + V82G + Y147T + Q154S)
ABE-823d	MSP823	heterodimer_(WT)+(TadA*7.10 + L36H + V82G + Y147T + Q154S + N157K)
ABE-824d	MSP824	heterodimer_(WT)+(TadA*7.10 + V82G + Y147D + F149Y + Q154S + D167N)
ABE-825d	MSP825	heterodimer_(WT)+(TadA*7.10 + L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N)
ABE-827d	MSP827	heterodimer_(WT)+(TadA*7.10 + L36H + I76Y + V82G + Y147T + Q154S + N157K)
ABE-828d	MSP828	heterodimer_(WT)+(TadA*7.10 + I76Y + V82G + Y147D + F149Y + Q154S + D167N)
ABE-829d	MSP829	heterodimer_(WT)+(TadA*7.10 + L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N)

In some embodiments, the base editor is a ninth generation ABE (ABE9). In some embodiments, the ABE9 contains a TadA*9 variant. ABE9 base editors include an adenosine deaminase variant comprising an amino acid sequence, which contains alterations relative to an ABE 7*10 reference sequence, as described herein. Exemplary ABE9 variants are listed in **Table 14A**.

Details of ABE9 base editors are described in International PCT Application No. PCT/2020/049975, which is incorporated herein by reference for its entirety, and are listed in **Table 14B**. In **Table 14B**, "monomer" indicates an ABE comprising a single TadA*7.10 comprising the indicated alterations and "heterodimer" indicates an ABE comprising a TadA*7.10 comprising the indicated alterations fused to an E. coli TadA adenosine deaminase.

Table 14B Adenosine Base Editor 9 (ABE9) Variants.

ABE9 Description	Alterations
ABE9.1_monomer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.2_monomer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.3_monomer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.4_monomer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.5_monomer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.6_monomer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.7_monomer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.8_monomer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.9_monomer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.10_monomer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.11_monomer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.12_monomer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.13_monomer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.14_monomer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.15_heterodimer	E25F, V82S, Y123H, T133K, Y147R, Q154R
ABE9.16_heterodimer	E25F, V82S, Y123H, Y147R, Q154R
ABE9.17_heterodimer	V82S, Y123H, P124W, Y147R, Q154R
ABE9.18_heterodimer	L51W, V82S, Y123H, C146R, Y147R, Q154R
ABE9.19_heterodimer	P54C, V82S, Y123H, Y147R, Q154R
ABE9.2_heterodimer	Y73S, V82S, Y123H, Y147R, Q154R
ABE9.21_heterodimer	N38G, V82T, Y123H, Y147R, Q154R
ABE9.22_heterodimer	R23H, V82S, Y123H, Y147R, Q154R
ABE9.23_heterodimer	R21N, V82S, Y123H, Y147R, Q154R
ABE9.24_heterodimer	V82S, Y123H, Y147R, Q154R, A158K
ABE9.25_heterodimer	N72K, V82S, Y123H, D139L, Y147R, Q154R,
ABE9.26_heterodimer	E25F, V82S, Y123H, D139M, Y147R, Q154R
ABE9.27_heterodimer	M70V, V82S, M94V, Y123H, Y147R, Q154R
ABE9.28_heterodimer	Q71M, V82S, Y123H, Y147R, Q154R
ABE9.29_monomer	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.30_monomer	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.31_monomer	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.32_monomer	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.33_monomer	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.34_monomer	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.35_monomer	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.36_monomer	I76Y_V82S_Y123H_D138M_Y147R_Q154R
ABE9.37_monomer	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.38_heterodimer	E25F_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.39_heterodimer	I76Y_V82T_Y123H_Y147R_Q154R
ABE9.40_heterodimer	N38G_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.41_heterodimer	N38G_I76Y_V82T_Y123H_Y147R_Q154R
ABE9.42_heterodimer	R23H_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.43_heterodimer	P54C_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.44_heterodimer	R21N_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.45_heterodimer	I76Y_V82S_Y123H_D138M_Y147R_Q154R

ABE9.46_heterodimer	Y72S_I76Y_V82S_Y123H_Y147R_Q154R
ABE9.47_monomer	N72K_V82S, Y123H, Y147R, Q154R
ABE9.48_monomer	Q71M_V82S, Y123H, Y147R, Q154R
ABE9.49_monomer	M70V,V82S, M94V, Y123H, Y147R, Q154R
ABE9.50_monomer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.51_monomer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.52_monomer	M70V,Q71M,N72K,V82S, Y123H, Y147R, Q154R
ABE9.53_heterodimer	N72K_V82S, Y123H, Y147R, Q154R
ABE9.54_heterodimer	Q71M_V82S, Y123H, Y147R, Q154R
ABE9.55_heterodimer	M70V,V82S, M94V, Y123H, Y147R, Q154R
ABE9.56_heterodimer	V82S, Y123H, T133K, Y147R, Q154R
ABE9.57_heterodimer	V82S, Y123H, T133K, Y147R, Q154R, A158K
ABE9.58_heterodimer	M70V, Q71M, N72K, V82S, Y123H, Y147R, Q154R

In some embodiments, the base editor is a fusion protein comprising a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9-derived domain) fused to a nucleobase editing domain (*e.g.*, all or a portion of a deaminase domain). In certain embodiments, the fusion proteins provided herein comprise one or more features that improve the base editing activity of the fusion proteins. For example, any of the fusion proteins provided herein may comprise a Cas9 domain that has reduced nuclease activity. In some embodiments, any of the fusion proteins provided herein may have a Cas9 domain that does not have nuclease activity (dCas9), or a Cas9 domain that cuts one strand of a duplexed DNA molecule, referred to as a Cas9 nickase (nCas9).

In some embodiments, the base editor further comprises a domain comprising all or a portion of a uracil glycosylase inhibitor (UGI). In some embodiments, the base editor comprises a domain comprising all or a portion of a uracil binding protein (UBP), such as a uracil DNA glycosylase (UDG). In some embodiments, the base editor comprises a domain comprising all or a portion of a nucleic acid polymerase. In some embodiments, a base editor can comprise as a domain all or a portion of a nucleic acid polymerase (NAP). For example, a base editor can comprise all or a portion of a eukaryotic NAP. In some embodiments, a NAP or portion thereof incorporated into a base editor is a DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor has translesion polymerase activity. In some embodiments, a NAP or portion thereof incorporated into a base editor is a translesion DNA polymerase. In some embodiments, a NAP or portion thereof incorporated into a base editor is a Rev7, Rev1 complex, polymerase iota, polymerase

kappa, or polymerase eta. In some embodiments, a NAP or portion thereof incorporated into a base editor is a eukaryotic polymerase alpha, beta, gamma, delta, epsilon, gamma, eta, iota, kappa, lambda, mu, or nu component. In some embodiments, a NAP or portion thereof incorporated into a base editor comprises an amino acid sequence that is at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to a nucleic acid polymerase (*e.g.*, a translesion DNA polymerase). In some embodiments, a nucleic acid polymerase or portion thereof incorporated into a base editor is a translesion DNA polymerase.

In some embodiments, a domain of the base editor can comprise multiple domains. For example, the base editor comprising a polynucleotide programmable nucleotide binding domain derived from Cas9 can comprise an REC lobe and an NUC lobe corresponding to the REC lobe and NUC lobe of a wild-type or natural Cas9. In another example, the base editor can comprise one or more of a RuvCI domain, BH domain, REC1 domain, REC2 domain, RuvCII domain, L1 domain, HNH domain, L2 domain, RuvCIII domain, WED domain, TOPO domain or CTD domain. In some embodiments, one or more domains of the base editor comprise a mutation (*e.g.*, substitution, insertion, deletion) relative to a wild-type version of a polypeptide comprising the domain. For example, an HNH domain of a polynucleotide programmable DNA binding domain can comprise an H840A substitution. In another example, a RuvCI domain of a polynucleotide programmable DNA binding domain can comprise a D10A substitution.

Different domains (*e.g.*, adjacent domains) of the base editor disclosed herein can be connected to each other with or without the use of one or more linker domains (*e.g.*, an XTEN linker domain). In some embodiments, a linker domain can be a bond (*e.g.*, covalent bond), chemical group, or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a first domain (*e.g.*, Cas9-derived domain) and a second domain (*e.g.*, an adenosine deaminase domain). In some embodiments, a linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-hetero atom bond, etc.). In certain embodiments, a linker is a carbon nitrogen bond of an amide linkage. In certain embodiments, a linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, a linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, etc.). In certain embodiments, a linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments, a linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, etc.). In some

embodiments, a linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, a linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, a linker comprises a polyethylene glycol moiety (PEG). In certain embodiments, a linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. A linker can include functionalized moieties to facilitate attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile can be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic acid editing protein. In some embodiments, a linker joins a dCas9 and a second domain (*e.g.*, UGI, etc.).

Linkers

In certain embodiments, linkers may be used to link any of the peptides or peptide domains as described herein. The linker may be as simple as a covalent bond, or it may be a polymeric linker many atoms in length. In certain embodiments, the linker is a polypeptide or based on amino acids. In other embodiments, the linker is not peptide-like. In certain embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, *etc.*). In certain embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In certain embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In certain embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In certain embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In certain embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In certain embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, the linker comprises a polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In certain embodiments, the linker comprises a peptide. In certain embodiments, the linker comprises an aryl or heteroaryl moiety. In certain embodiments, the linker is based on a phenyl ring. The linker may include functionalized moieties to facilitate

attachment of a nucleophile (*e.g.*, thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

Typically, a linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, a linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, a linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, a linker is 2-100 amino acids in length, for example, 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, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. In some embodiments, the linker is about 3 to about 104 (*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, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100) amino acids in length. Longer or shorter linkers are also contemplated.

In some embodiments, any of the fusion proteins provided herein, comprise a adenosine deaminase and a Cas9 domain that are fused to each other via a linker. Various linker lengths and flexibilities between the adenosine deaminase and the Cas9 domain can be employed (*e.g.*, ranging from very flexible linkers of the form (GGGS)n (SEQ ID NO: 246), (GGGGS)n (SEQ ID NO: 247), and (G)n to more rigid linkers of the form (EAAAK)n (SEQ ID NO: 248), (SGGS)n (SEQ ID NO: 355), SGSETPGTSESATPES (SEQ ID NO: 249) (see, *e.g.*, Guilinger JP, *et al.* Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 2014; 32(6): 577-82; the entire contents are incorporated herein by reference) and $(XP)_n$) in order to achieve the optimal length for activity for the adenosine deaminase nucleobase editor. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, the linker comprises a $(GGS)_n$ motif, wherein n is 1, 3, or 7. In some embodiments, adenosine deaminase and the Cas9 domain of any of the fusion proteins provided herein are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES, which can also be referred to as the XTEN linker. In some embodiments, a linker comprises a plurality of proline residues and is 5-21, 5-14, 5-9, 5-7 amino acids in length, e.g., PAPAP (SEQ ID NO: 363), PAPAPA (SEQ ID NO: 364), PAPAPAP (SEQ ID NO: 365), PAPAPAPA (SEQ ID NO: 366), P(AP)4 (SEQ ID NO: 367), P(AP)7 (SEQ ID NO: 368), P(AP)10 (SEQ ID NO: 369) (see, *e.g.*, Tan J, Zhang F,

Karcher D, Bock R. Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat Commun. 2019 Jan 25;10(1):439; the entire contents are incorporated herein by reference). Such proline-rich linkers are also termed "rigid" linkers. In some embodiments, adenosine deaminase and the Cas9 domain of any of the fusion proteins provided herein are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 249), which can also be referred to as the XTEN linker.

In some embodiments, the domains of the base editor are fused via a linker that comprises the amino acid sequence of:
SGGSSGSETPGTSESATPESSGGS (SEQ ID NO: 356),
SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 357), or
GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS
EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 358).

In some embodiments, domains of the base editor are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 249), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGS. In some embodiments, the linker is 24 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 359). In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS (SEQ ID NO: 360). In some embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence:
SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSG
GS (SEQ ID NO: 361). In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence:
PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG
TSTEPSEGSAPGTSESATPESGPGSEPATS (SEQ ID NO: 362).

In another embodiment, the base editor system comprises a component (protein) that interacts non-covalently with a deaminase (DNA deaminase), *e.g.*, an adenosine, and transiently attracts the adenosine deaminase to the target nucleobase in a target polynucleotide sequence for specific editing, with minimal or reduced bystander or target-adjacent effects. Such a non-covalent system and method involving deaminase-interacting

proteins serves to attract a DNA deaminase to a particular genomic target nucleobase and decouples the events of on-target and target-adjacent editing, thus enhancing the achievement of more precise single base substitution mutations. In an embodiment, the deaminase-interacting protein binds to the deaminase (*e.g.*, adenosine deaminase) without blocking or interfering with the active (catalytic) site of the deaminase from engaging the target nucleobase (*e.g.*, adenosine). Such as system, termed "MagnEdit," involves interacting proteins tethered to a Cas9 and gRNA complex and can attract a co-expressed adenosine (either exogenous or endogenous) to edit a specific genomic target site, and is described in McCann, J. *et al.*, 2020, "MagnEdit – interacting factors that recruit DNA-editing enzymes to single base targets," Life-Science-Alliance, Vol. 3, No. 4 (e201900606), (doi 10.26508/lsa.201900606), the contents of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (*e.g.*, TadA*8) as described herein.

In another embodiment, a system called "Suntag," involves non-covalently interacting components used for recruiting protein (*e.g.*, adenosine deaminase) components, or multiple copies thereof, of base editors to polynucleotide target sites to achieve base editing at the site with reduced adjacent target editing, for example, as described in Tanenbaum, M.E. *et al.*, "A protein tagging system for signal amplification in gene expression and fluorescence imaging," *Cell*. 2014 October 23; 159(3): 635–646. doi:10.1016/j.cell.2014.09.039; and in Huang, Y.-H. *et al.*, 2017, "DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A," *Genome Biol* 18: 176. doi:10.1186/s13059-017-1306-z, the contents of each of which are incorporated by reference herein in their entirety. In an embodiment, the DNA deaminase is an adenosine deaminase variant (*e.g.*, TadA*8) as described herein.

Nucleic acid programmable DNA binding proteins with guide RNAs

Some aspects of this disclosure provide complexes comprising any of the fusion proteins provided herein, and a guide RNA bound to a nucleic acid programmable DNA binding protein (napDNAbp)domain (*e.g.*, a Cas9 (*e.g.*, a dCas9, a nuclease active Cas9, or a Cas9 nickase)) of the fusion protein. These complexes are also termed ribonucleoproteins (RNPs). In some embodiments, the guide nucleic acid (*e.g.*, guide RNA) is from 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 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, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is an RNA sequence. In some embodiments, the target sequence is a sequence in the genome of a bacteria, yeast, fungi, insect, plant, or animal. In some embodiments, the target sequence is a sequence in the genome of a human. In some embodiments, the 3' end of the target sequence is immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is immediately adjacent to a non-canonical PAM sequence (*e.g.*, a sequence listed in **Table 2** or 5'-NAA-3'). In some embodiments, the guide nucleic acid (*e.g.*, guide RNA) is complementary to a sequence in a G6PC allele bearing GSD1a targetable mutations.

Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an AGC, GAG, TTT, GTG, or CAA sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, or 5' (TTTV) sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an *e.g.*, TTN, DTTN, GTTN, ATTN, ATTC, DTTNT, WTTN, HATY, TTTN, TTTV, TTTC, TG, RTR, or YTN PAM site.

It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might differ, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues.

It will be apparent to those of skill in the art that in order to target any of the fusion proteins disclosed herein, to a target site, *e.g.*, a site comprising a mutation to be edited, it is typically necessary to co-express the fusion protein together with a guide RNA. As explained

in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for napDNAbp (*e.g.*, Cas9) binding, and a guide sequence, which confers sequence specificity to the napDNAbp:nucleic acid editing enzyme/domain fusion protein. Alternatively, the guide RNA and tracrRNA may be provided separately, as two nucleic acid molecules. In some embodiments, the guide RNA comprises a structure, wherein the guide sequence comprises a sequence that is complementary to the target sequence. The guide sequence is typically 20 nucleotides long. The sequences of suitable guide RNAs for targeting napDNAbp:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein.

Distinct portions of sgRNA are predicted to form various features that interact with Cas9 (*e.g.*, SpyCas9) and/or the DNA target. Six conserved modules have been identified within native crRNA:tracrRNA duplexes and single guide RNAs (sgRNAs) that direct Cas9 endonuclease activity (*see* Briner *et al.*, Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339). The six modules include the spacer responsible for DNA targeting, the upper stem, bulge, lower stem formed by the CRISPR repeat:tracrRNA duplex, the nexus, and hairpins from the 3' end of the tracrRNA. The upper and lower stems interact with Cas9 mainly through sequence-independent interactions with the phosphate backbone. In some embodiments, the upper stem is dispensable. In some embodiments, the conserved uracil nucleotide sequence at the base of the lower stem is dispensable. The bulge participates in specific side-chain interactions with the Rec1 domain of Cas9. The nucleobase of U44 interacts with the side chains of Tyr 325 and His 328, while G43 interacts with Tyr 329. The nexus forms the core of the sgRNA:Cas9 interactions and lies at the intersection between the sgRNA and both Cas9 and the target DNA. The nucleobases of A51 and A52 interact with the side chain of Phe 1105; U56 interacts with Arg 457 and Asn 459; the nucleobase of U59 inserts into a hydrophobic pocket defined by side chains of Arg 74, Asn 77, Pro 475, Leu 455, Phe 446, and Ile 448; C60 interacts with Leu 455, Ala 456, and Asn 459, and C61 interacts with the side chain of Arg 70, which in turn interacts with C15. In some embodiments, one or more of these

mutations are made in the bulge and/or the nexus of a sgRNA for a Cas9 (*e.g.*, spyCas9) to optimize sgRNA:Cas9 interactions.

Moreover, the tracrRNA nexus and hairpins are critical for Cas9 pairing and can be swapped to cross orthogonality barriers separating disparate Cas9 proteins, which is instrumental for further harnessing of orthogonal Cas9 proteins. In some embodiments, the nexus and hairpins are swapped to target orthogonal Cas9 proteins. In some embodiments, a sgRNA is dispensed of the upper stem, hairpin 1, and/or the sequence flexibility of the lower stem to design a guide RNA that is more compact and conformationally stable. In some embodiments, the modules are modified to optimize multiplex editing using a single Cas9 with various chimeric guides or by concurrently using orthogonal systems with different combinations of chimeric sgRNAs. Details regarding guide functional modules and methods thereof are described, for example, in Briner *et al.*, Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality Mol Cell. 2014 Oct 23;56(2):333-339, the contents of which is incorporated by reference herein in its entirety. The domains of the base editor disclosed herein can be arranged in any order. Non-limiting examples of a base editor comprising a fusion protein comprising *e.g.*, a polynucleotide-programmable nucleotide-binding domain (*e.g.*, Cas9) and a deaminase domain (*e.g.*, adenosine deaminase) can be arranged as follows:

NH_2-[nucleobase editing domain]-Linker1-[nucleobase editing domain]-COOH;

NH_2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;

NH_2-[deaminase]-Linker1-[nucleobase editing domain]-Linker2-[UGI]-COOH;

NH_2-[deaminase]-Linker1-[nucleobase editing domain]-COOH;

NH_2-[adenosine deaminase]-Linker1-[nucleobase editing domain]-COOH;

NH_2-[nucleobase editing domain]-[deaminase]-COOH;

NH_2-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;

NH_2-[deaminase]-[inosine BER inhibitor]-[nucleobase editing domain]-COOH;

NH_2-[inosine BER inhibitor]-[deaminase]-[nucleobase editing domain]-COOH;

NH_2-[nucleobase editing domain]-[deaminase]-[inosine BER inhibitor]-COOH;

NH_2-[nucleobase editing domain]-[inosine BER inhibitor]-[deaminase]-COOH;

NH_2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH;

NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH;
NH2-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
NH2-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
NH2-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-[inosine BER inhibitor]-COOH;
NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-Linker2-[nucleobase editing domain]-COOH;
NH2-[inosine BER inhibitor]-[nucleobase editing domain]-Linker1-[deaminase]-[nucleobase editing domain]-COOH;
NH2-[inosine BER inhibitor]-[nucleobase editing domain]-[deaminase]-Linker2-[nucleobase editing domain]-COOH; or
NH2-[inosine BER inhibitor]NH2-[nucleobase editing domain]-[deaminase]-[nucleobase editing domain]-COOH.

In some embodiments, the base editing fusion proteins provided herein need to be positioned at a precise location, for example, where a target base is placed within a defined region (*e.g.*, a "deamination window"). In some embodiments, a target can be within a 4-base region. In some embodiments, such a defined target region can be approximately 15 bases upstream of the PAM. See Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference.

A defined target region can be a deamination window. A deamination window can be the defined region in which a base editor acts upon and deaminates a target nucleotide. In some embodiments, the deamination window is within a 2, 3, 4, 5, 6, 7, 8, 9, or 10 base regions. In some embodiments, the deamination window is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bases upstream of the PAM.

The base editors of the present disclosure can comprise any domain, feature or amino acid sequence which facilitates the editing of a target polynucleotide sequence. For example, in some embodiments, the base editor comprises a nuclear localization sequence (NLS). In some embodiments, an NLS of the base editor is localized between a deaminase domain and a napDNAbp domain. In some embodiments, an NLS of the base editor is localized C-terminal to a napDNAbp domain.

Non-limiting examples of protein domains which can be included in the fusion protein include a deaminase domain (*e.g.*, adenosine deaminase), a uracil glycosylase inhibitor (UGI) domain, epitope tags, reporter gene sequences, and/or protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, gene silencing activity, chromatin modifying activity, epigenetic modifying activity, histone modification activity, RNA cleavage activity, and nucleic acid binding activity. Additional domains can be a heterologous functional domain. Such heterologous functional domains can confer a function activity, such as modification of a target polypeptide associated with target DNA (*e.g.*, a histone, a DNA binding protein, etc.), leading to, for example, histone methylation, histone acetylation, histone ubiquitination, and the like. Other functions and/or activities conferred can include transposase activity, integrase activity, recombinase activity, ligase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylation activity, deSUMOylation activity, or any combination of the above.

Other functions conferred can include methyltransferase activity, demethylase activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, remodeling activity, protease activity, oxidoreductase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, synthase activity, synthetase activity, and demyristoylation activity, or any combination thereof.

A domain may be detected or labeled with an epitope tag, a reporter protein, other binding domains. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-5-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). Additional protein sequences can include amino acid sequences that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions.

Methods of using fusion proteins comprising adenosine deaminase and a Cas9 domain

Some aspects of this disclosure provide methods of using the fusion proteins, or complexes provided herein. For example, some aspects of this disclosure provide methods comprising contacting a DNA molecule with any of the fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is not immediately adjacent to a canonical PAM sequence (NGG). In some embodiments, the 3' end of the target sequence is immediately adjacent to an AGC, GAG, TTT, GTG, or CAA sequence. In some embodiments, the 3' end of the target sequence is immediately adjacent to an NGA, NGCG, NGN, NNGRRT, NNNRRT, NGCG, NGCN, NGTN, NGTN, NGTN, or 5' (TTTV) sequence.

Base Editor Efficiency

In some embodiments, the nucleobase editing proteins provided herein can be validated for gene editing-based human therapeutics *in vitro*. It will be understood by the skilled artisan that the nucleobase editing proteins provided herein, *e.g.*, the fusion proteins comprising a polynucleotide programmable nucleotide binding domain (*e.g.*, Cas9) and a

nucleobase editing domain (*e.g.*, an adenosine deaminase domain) can be used to edit a nucleotide from A to G or C to T .

As most of the known genetic variations associated with human disease are point mutations, methods that can more efficiently and cleanly make precise point mutations are needed. Base editing systems as provided herein provide a new way to provide genome editing without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions.

In some embodiments, the present disclosure provides base editors that efficiently generate an intended mutation in a polynucleotide sequence without generating a significant number of unintended mutations, such as unintended point mutations. In some embodiments, an intended mutation is a mutation that is generated by a specific base editor (*e.g.*, adenosine base editor) bound to a guide polynucleotide (*e.g.*, gRNA), specifically designed to generate the intended mutation. In some embodiments, the intended mutation is in a gene associated with a target antigen associated with a disease or disorder, *e.g.*, a mutation in the *G6PC* gene associated with GSD1a. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation (*e.g.*, SNP) in a gene associated with a target antigen associated with a disease or disorder, *e.g.*, a mutation in the *G6PC* gene associated with GSD1a. In some embodiments, the intended mutation is an adenine (A) to guanine (G) point mutation within the coding region or non-coding region of a gene (*e.g.*, regulatory region or element).

The base editors as described herein advantageously modify a specific nucleotide base encoding a protein without generating a significant proportion of indels. An "indel", as used herein, refers to the insertion or deletion of a nucleotide base within a nucleic acid. Such insertions or deletions can lead to frame shift mutations within a coding region of a gene. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.* mutate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In some embodiments, it is desirable to generate base editors that efficiently modify (*e.g.* mutate or methylate) a specific nucleotide within a nucleic acid, without generating a large number of insertions or deletions (*i.e.*, indels) in the nucleic acid. In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, methylations) versus indels. In certain embodiments, any of the base editors provided herein can generate a greater proportion of intended modifications (*e.g.*, mutations) versus indels.

In some embodiments, the base editors provided herein are capable of generating a ratio of intended mutations to indels (*i.e.*, intended point mutations:unintended point mutations) that is greater than 1:1. In some embodiments, the base editors provided herein are capable of generating a ratio of intended mutations to indels that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 200:1, at least 300:1, at least 400:1, at least 500:1, at least 600:1, at least 700:1, at least 800:1, at least 900:1, or at least 1000:1, or more. The number of intended mutations and indels may be determined using any suitable method.

In some embodiments, the base editors provided herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor. In some embodiments, any of the base editors provided herein can limit the formation of indels at a region of a nucleic acid to less than 1%, less than 1.5%, less than 2%, less than 2.5%, less than 3%, less than 3.5%, less than 4%, less than 4.5%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, less than 10%, less than 12%, less than 15%, or less than 20%. The number of indels formed at a nucleic acid region may depend on the amount of time a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, a number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) to a base editor.

Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation in a nucleic acid (*e.g.* a nucleic acid within a genome of a subject) without generating a considerable number of unintended mutations (*e.g.*, spurious off-target editing or bystander editing). In some embodiments, an intended mutation is a mutation that is generated by a specific base editor bound to a gRNA, specifically designed to generate the intended mutation. In some embodiments, the intended mutation is a mutation that generates a stop codon, for example, a premature stop codon within the coding region of a gene. In some embodiments, the intended mutation is a mutation that eliminates a stop codon. In some embodiments, the

intended mutation is a mutation that alters the splicing of a gene. In some embodiments, the intended mutation is a mutation that alters the regulatory sequence of a gene (*e.g.*, a gene promotor or gene repressor).

In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations (*e.g.*, intended mutations:unintended mutations) that is greater than 1:1. In some embodiments, any of the base editors provided herein are capable of generating a ratio of intended mutations to unintended mutations that is at least 1.5:1, at least 2:1, at least 2.5:1, at least 3:1, at least 3.5:1, at least 4:1, at least 4.5:1, at least 5:1, at least 5.5:1, at least 6:1, at least 6.5:1, at least 7:1, at least 7.5:1, at least 8:1, at least 10:1, at least 12:1, at least 15:1, at least 20:1, at least 25:1, at least 30:1, at least 40:1, at least 50:1, at least 100:1, at least 150:1, at least 200:1, at least 250:1, at least 500:1, or at least 1000:1, or more. It should be appreciated that the characteristics of the base editors described herein, may be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Base editing is often referred to as a "modification", such as, a genetic modification, a gene modification and modification of the nucleic acid sequence and is clearly understandable based on the context that the modification is a base editing modification. A base editing modification is therefore a modification at the nucleotide base level, for example as a result of the deaminase activity discussed throughout the disclosure, which then results in a change in the gene sequence, and may affect the gene product. In essence therefore, the gene editing modification described herein may result in a modification of the gene, structurally and/or functionally, wherein the expression of the gene product may be modified, for example, the expression of the gene is knocked out; or conversely, enhanced, or, in some circumstances, the gene function or activity may be modified. Using the methods disclosed herein, a base editing efficiency may be determined as the knockdown efficiency of the gene in which the base editing is performed, wherein the base editing is intended to knockdown the expression of the gene. A knockdown level may be validated quantitatively by determining the expression level by any detection assay, such as assay for protein expression level, for example, by flow cytometry; assay for detecting RNA expression such as quantitative RT-PCR, northern blot analysis, or any other suitable assay such as pyrosequencing; and may be validated qualitatively by nucleotide sequencing reactions.

In some embodiments, any of base editor systems provided herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than

17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% indel formation in the target polynucleotide sequence.

In some embodiments, targeted modifications, *e.g.*, single base editing, are used simultaneously to target at least 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 or 50 different endogenous sequences for base editing with different guide RNAs. In some embodiments, targeted modifications, *e.g.* single base editing, are used to sequentially target at least 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, or more different endogenous gene sequences for base editing with different guide RNAs.

Some aspects of the disclosure are based on the recognition that any of the base editors provided herein are capable of efficiently generating an intended mutation, such as a point mutation, in a nucleic acid (*e.g.*, a nucleic acid within a genome of a subject) without generating a significant number of unintended mutations, such as unintended point mutations (*i.e.*, mutation of bystanders). In some embodiments, any of the base editors provided herein are capable of generating at least 0.01% of intended mutations (*i.e.*, at least 0.01% base editing efficiency). In some embodiments, any of the base editors provided herein are capable of generating at least 0.01%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% of intended mutations.

In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than

0.02%, or less than 0.01% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein result in less than 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein result in at most 0.8% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein result in less than 0.3% indel formation in the target polynucleotide sequence. In some embodiments, any of base editor systems comprising one of the ABE base editor variants described results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising one of ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein results in lower indel formation in the target polynucleotide sequence compared to a base editor system comprising an ABE7.10.

In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein has reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising one of the ABE7 base editors. In some embodiments, a base editor system comprising one of the ABE base editor variants described herein has at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% reduction in indel frequency compared to a base editor system comprising an ABE7.10.

Provided herein are adenosine deaminase variants (*e.g.*, TadA variants) that have increased efficiency and specificity. In particular, the adenosine deaminase variants described herein are more likely to edit a desired base within a polynucleotide, and are less likely to edit bases that are not intended to be altered (*e.g.*, "bystanders").

In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced bystander editing or mutations. In some embodiments, an unintended editing or mutation is a bystander mutation or bystander editing, for example, base editing of a target base (*e.g.*, A or C) in an unintended or non-target position in a target window of a target nucleotide sequence. In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced bystander editing or mutations compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced bystander editing or mutations by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced bystander editing or mutations by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced spurious editing. In some embodiments, an unintended editing or mutation is a spurious mutation or spurious editing, for example, non-specific editing or guide independent editing of a target base (*e.g.*, A or C) in an unintended or non-target region of the genome. In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced spurious editing compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced spurious editing by at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10. In some

embodiments, any of the base editing system comprising one of the ABE base editor variants described herein has reduced spurious editing by at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold compared to a base editor system comprising an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% base editing efficiency. In some embodiments, the base editing efficiency may be measured by calculating the percentage of edited nucleobases in a population of cells. In some embodiments, any of the ABE base editor variants described herein have base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases in a population of cells.

In some embodiments, any of the ABE base editor variants described herein has higher base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE base editor variants described herein have at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% on-target base editing efficiency. In some embodiments, any of the ABE base editor variants described herein have on-target base editing efficiency of at least 0.01%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited target nucleobases in a population of cells.

In some embodiments, any of the ABE base editor variants described herein has higher on-target base editing efficiency compared to the ABE7 base editors. In some embodiments, any of the ABE base editor variants described herein have at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300%, at least 310%, at least 320%, at least 330%, at least 340%, at least

350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher on-target base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

In some embodiments, any of the ABE base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target base editing efficiency compared to an ABE7 base editor, *e.g.*, ABE7.10.

The ABE base editor variants described herein may be delivered to a host cell via a plasmid, a vector, a LNP complex, or an mRNA. In some embodiments, any of the ABE base editor variants described herein is delivered to a host cell as an mRNA. In some embodiments, an ABE base editor delivered via a nucleic acid based delivery system, *e.g.*, an mRNA, has on-target editing efficiency of at least at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% as measured by edited nucleobases. In some embodiments, an ABE base editor delivered by an mRNA system has higher base editing efficiency compared to an ABE base editor delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at

least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 3.6 fold, at least 3.7 fold, at least 3.8 fold, at least 3.9 fold, at least 4.0 fold, at least 4.1 fold, at least 4.2 fold, at least 4.3 fold, at least 4.4 fold, at least 4.5 fold, at least 4.6 fold, at least 4.7 fold, at least 4.8 fold, at least 4.9 fold, or at least 5.0 fold higher on-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of base editor systems comprising one of the ABE base editor variants described herein result in less than 50%, less than 40%, less than 30%, less than 20%, less than 19%, less than 18%, less than 17%, less than 16%, less than 15%, less than 14%, less than 13%, less than 12%, less than 11%, less than 10%, less than 9%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.9%, less than 0.8%, less than 0.7%, less than 0.6%, less than 0.5%, less than 0.4%, less than 0.3%, less than 0.2%, less than 0.1%, less than 0.09%, less than 0.08%, less than 0.07%, less than 0.06%, less than 0.05%, less than 0.04%, less than 0.03%, less than 0.02%, or less than 0.01% off-target editing in the target polynucleotide sequence.

In some embodiments, any of the ABE base editor variants described herein has lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0

fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, or at least 3.0 fold lower guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least about 2.2 fold decrease in guided off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system.

In some embodiments, any of the ABE base editor variants described herein has lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, any of the ABE base editor variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 5.0 fold, at least 10.0 fold, at least 20.0 fold, at least 50.0 fold, at least 70.0 fold, at least 100.0 fold, at least 120.0 fold, at least 130.0 fold, or at least 150.0 fold lower guide-independent off-target editing efficiency when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE base editor variants described herein has 134.0 fold decrease in guide-independent off-target editing efficiency (*e.g.*, spurious RNA deamination) when delivered by an mRNA system compared to when delivered by a plasmid or vector system. In some embodiments, ABE base editor variants described herein does not increase guide-independent mutation rates across the genome.

In some embodiments, a single gene delivery event (*e.g.*, by transduction, transfection, electroporation or any other method) can be used to target base editing of 5 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 6 sequences within a cell's genome. In some embodiments, a

single gene delivery event can be used to target base editing of 7 sequences within a cell's genome. In some embodiments, a single electroporation event can be used to target base editing of 8 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 9 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 10 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 20 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 30 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 40 sequences within a cell's genome. In some embodiments, a single gene delivery event can be used to target base editing of 50 sequences within a cell's genome.

In some embodiments, the method described herein, for example, the base editing methods has minimum to no off-target effects.

In some embodiments, the base editing method described herein results in at least 50% of a cell population that have been successfully edited (*i.e.*, cells that have been successfully engineered). In some embodiments, the base editing method described herein results in at least 55% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 60% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 65% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 70% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 75% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 80% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 85% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 90% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in at least 95% of a cell population that have been successfully edited. In some embodiments, the base editing method described herein results in about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% of a cell population that have been successfully edited.

In some embodiments, the live cell recovery following a base editing intervention is greater than at least 60%, 70%, 80%, 90% of the starting cell population at the time of the base editing event. In some embodiments, the live cell recovery as described above is about 70%. In some embodiments, the live cell recovery as described above is about 75%. In some embodiments, the live cell recovery as described above is about 80%. In some embodiments, the live cell recovery as described above is about 85%. In some embodiments, the live cell recovery as described above is about 90%, or about 91%, 92%, 93%, 94% 95%, 96%, 97%, 98%, or 99%, or 100% of the cells in the population at the time of the base editing event.

In some embodiments the engineered cell population can be further expanded in vitro by about 2 fold, about 3-fold, about 4-fold, about 5-fold, about 6-fold, about 7-fold, about 8-fold, about 9-fold, about 10-fold, about 15-fold, about 20-fold, about 25-fold, about 30-fold, about 35-fold, about 40-fold, about 45-fold, about 50-fold, or about 100-fold.

The number of intended mutations and indels can be determined using any suitable method, for example, as described in International PCT Application Nos. PCT/2017/045381 (WO2018/027078) and PCT/US2016/058344 (WO2017/070632); Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017); the entire contents of which are hereby incorporated by reference.

In some embodiments, to calculate indel frequencies, sequencing reads are scanned for exact matches to two 10-bp sequences that flank both sides of a window in which indels can occur. If no exact matches are located, the read is excluded from analysis. If the length of this indel window exactly matches the reference sequence the read is classified as not containing an indel. If the indel window is two or more bases longer or shorter than the reference sequence, then the sequencing read is classified as an insertion or deletion, respectively. In some embodiments, the base editors provided herein can limit formation of indels in a region of a nucleic acid. In some embodiments, the region is at a nucleotide targeted by a base editor or a region within 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nucleotide targeted by a base editor.

The number of indels formed at a target nucleotide region can depend on the amount of time a nucleic acid (*e.g.*, a nucleic acid within the genome of a cell) is exposed to a base editor. In some embodiments, the number or proportion of indels is determined after at least 1 hour, at least 2 hours, at least 6 hours, at least 12 hours, at least 24 hours, at least 36 hours, at least 48 hours, at least 3 days, at least 4 days, at least 5 days, at least 7 days, at least 10 days, or at least 14 days of exposing the target nucleotide sequence (*e.g.*, a nucleic acid within the genome of a cell) to a base editor. It should be appreciated that the characteristics of the base editors as described herein can be applied to any of the fusion proteins, or methods of using the fusion proteins provided herein.

Details of base editor efficiency are described in International PCT Application Nos. PCT/2017/045381 (WO 2018/027078) and PCT/US2016/058344 (WO 2017/070632), each of which is incorporated herein by reference for its entirety. Also see Komor, A.C., *et al.*, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., *et al.*, "Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); and Komor, A.C., *et al.*, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), the entire contents of which are hereby incorporated by reference. In some embodiments, editing of a plurality of nucleobase pairs in one or more genes using the methods provided herein results in formation of at least one intended mutation. In some embodiments, said formation of said at least one intended mutation results in the disruption the normal function of a gene. In some embodiments, said formation of said at least one intended mutation results decreases or eliminates the expression of a protein encoded by a gene. It should be appreciated that multiplex editing can be accomplished using any method or combination of methods provided herein.

Multiplex Editing

In some embodiments, the base editor system provided herein is capable of multiplex editing of a plurality of nucleobase pairs in one or more genes. In some embodiments, the plurality of nucleobase pairs is located in the same gene. In some embodiments, the plurality of nucleobase pairs is located in one or more gene, wherein at least one gene is located in a different locus. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more base

editor systems. In some embodiments, the multiplex editing can comprise one or more base editor systems with a single guide polynucleotide. In some embodiments, the multiplex editing can comprise one or more base editor systems with a plurality of guide polynucleotides. In some embodiments, the multiplex editing can comprise one or more guide polynucleotides with a single base editor system. In some embodiments, the multiplex editing can comprise at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, multiplex editing can comprise at least one guide polynucleotide that requires a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the multiplex editing can comprise a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any combination of methods using any base editor provided herein. It should also be appreciated that the multiplex editing using any of the base editors as described herein can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the plurality of nucleobase pairs are in one more genes. In some embodiments, the plurality of nucleobase pairs is in the same gene. In some embodiments, at least one gene in the one more genes is located in a different locus.

In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein coding region. In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein non-coding region. In some embodiments, the editing is editing of the plurality of nucleobase pairs in at least one protein coding region and at least one protein non-coding region.

In some embodiments, the editing is in conjunction with one or more guide polynucleotides. In some embodiments, the base editor system can comprise one or more base editor system. In some embodiments, the base editor system can comprise one or more base editor systems in conjunction with a single guide polynucleotide. In some embodiments, the base editor system can comprise one or more base editor system in conjunction with a plurality of guide polynucleotides. In some embodiments, the editing is in conjunction with one or more guide polynucleotide with a single base editor system. In some embodiments, the editing is in conjunction with at least one guide polynucleotide that does

not require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the editing is in conjunction with at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. In some embodiments, the editing is in conjunction with a mix of at least one guide polynucleotide that does not require a PAM sequence to target binding to a target polynucleotide sequence and at least one guide polynucleotide that require a PAM sequence to target binding to a target polynucleotide sequence. It should be appreciated that the characteristics of the multiplex editing using any of the base editors as described herein can be applied to any of combination of the methods of using any of the base editors provided herein. It should also be appreciated that the editing can comprise a sequential editing of a plurality of nucleobase pairs.

In some embodiments, the base editor system capable of multiplex editing of a plurality of nucleobase pairs in one or more genes comprises one of the ABE base editor variants described herein. In some embodiments, the base editor system capable of multiplex editing of a plurality of nucleobase pairs in one or more genes comprises one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE base editor variants described herein has higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE base editor variants described herein has at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, at least 100%, at least 105%, at least 110%, at least 115%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 155%, at least 160%, at least 165%, at least 170%, at least 175%, at least 180%, at least 185%, at least 190%, at least 195%, at least 200%, at least 210%, at least 220%, at least 230%, at least 240%, at least 250%, at least 260%, at least 270%, at least 280%, at least 290%, at least 300% higher, at least 310%, at least 320%, at least 330%, at least 340%, at least 350%, at least 360%, at least 370%, at least 380%, at least 390%, at least 400%, at least 450%, or at least 500% higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors. In some embodiments, the base editor system capable of multiplex editing comprising one of the ABE base editor

variants described herein has at least 1.1 fold, at least 1.2 fold, at least 1.3 fold, at least 1.4 fold, at least 1.5 fold, at least 1.6 fold, at least 1.7 fold, at least 1.8 fold, at least 1.9 fold, at least 2.0 fold, at least 2.1 fold, at least 2.2 fold, at least 2.3 fold, at least 2.4 fold, at least 2.5 fold, at least 2.6 fold, at least 2.7 fold, at least 2.8 fold, at least 2.9 fold, at least 3.0 fold, at least 3.1 fold, at least 3.2 fold, at least 3.3 fold, at least 3.4 fold, at least 3.5 fold, at least 4.0 fold, at least 4.5 fold, at least 5.0 fold, at least 5.5 fold, or at least 6.0 fold higher multiplex editing efficiency compared the base editor system capable of multiplex editing comprising one of ABE7 base editors.

DELIVERY SYSTEM

Nucleic Acid-Based Delivery of Nucleobase Editors and gRNAs

Nucleic acids encoding an adenosine deaminase nucleobase editor according to the present disclosure can be administered to subjects or delivered into cells *in vitro, ex vivo,* or *in vivo* by art-known methods or as described herein. For example, adenosine deaminase nucleobase editors can be delivered by, *e.g.*, vectors (*e.g.*, viral or non-viral vectors), non-vector based methods (*e.g.*, using naked DNA, DNA complexes, lipid nanoparticles), or a combination thereof.

Nucleic acids encoding adenosine deaminase nucleobase editors can be delivered directly to cells (*e.g.*, hematopoietic cells or their progenitors, hematopoietic stem cells, and/or induced pluripotent stem cells) as naked DNA or RNA, for instance by means of transfection or electroporation, or can be conjugated to molecules (*e.g.*, N-acetylgalactosamine) promoting uptake by the target cells. Nucleic acid vectors, such as the vectors described herein can also be used.

Nucleic acid vectors can comprise one or more sequences encoding a domain of a fusion protein described herein. A vector can also comprise a sequence encoding a signal peptide (*e.g.*, for nuclear localization, nucleolar localization, or mitochondrial localization), associated with (*e.g.*, inserted into or fused to) a sequence coding for a protein. As one example, a nucleic acid vectors can include a Cas9 coding sequence that includes one or more nuclear localization sequences (*e.g.*, a nuclear localization sequence from SV40), and an adenosine deaminase variant (*e.g.*, TadA variant).

The nucleic acid vector can also include any suitable number of regulatory/control elements, *e.g.*, promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, or internal ribosome entry sites (IRES). These elements are well known in the art.

Nucleic acid vectors according to this disclosure include recombinant viral vectors. Exemplary viral vectors are set forth herein. Other viral vectors known in the art can also be used. In addition, viral particles can be used to deliver base editing system components in nucleic acid and/or peptide form. For example, "empty" viral particles can be assembled to contain any suitable cargo. Viral vectors and viral particles can also be engineered to incorporate targeting ligands to alter target tissue specificity.

In addition to viral vectors, non-viral vectors can be used to deliver nucleic acids encoding genome editing systems according to the present disclosure. One important category of non-viral nucleic acid vectors are nanoparticles, which can be organic or inorganic. Nanoparticles are well known in the art. Any suitable nanoparticle design can be used to deliver genome editing system components or nucleic acids encoding such components. For instance, organic (*e.g.* lipid and/or polymer) nanoparticles can be suitable for use as delivery vehicles in certain embodiments of this disclosure.

Nucleic Acid-Based Delivery of Base Editor Systems

Nucleic acid molecules encoding a base editor system according to the present disclosure can be administered to subjects or delivered into cells *in vitro* or *in vivo* by art-known methods or as described herein. For example, a base editor system comprising a deaminase (*e.g.*, cytidine or adenine deaminase) can be delivered by vectors (*e.g.*, viral or non-viral vectors), or by naked DNA, DNA complexes, lipid nanoparticles, or a combination of the aforementioned compositions.

Nanoparticles, which can be organic or inorganic, are useful for delivering a base editor system or component thereof. Nanoparticles are well known in the art and any suitable nanoparticle can be used to deliver a base editor system or component thereof, or a nucleic acid molecule encoding such components. In one example, organic (*e.g.* lipid and/or polymer) nanoparticles are suitable for use as delivery vehicles in certain embodiments of this disclosure.

Exemplary lipids for use in nanoparticle formulations, and/or gene transfer are shown in **Table 15** (below).

Table 15

Lipids Used for Gene Transfer		
Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleyloxy)prophyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Dioctadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleyloxy-N-[2(sperminecarboxamido-ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diodeoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Dioctadecylamidoglicylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammoniun bromide	CLIP-6	Cationic
Ethyldimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearyloxy-N,N-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylpho sphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl] imidazolinium chloride	DOTIM	Cationic
N1 -Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3-dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl-methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 16 lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 16

Polymers Used for Gene Transfer	
Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis (succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	DTBP
Poly(ethylene imine)biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(N-vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amidoethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β-aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α-[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(N-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(N-2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

Table 17 summarizes delivery methods for a polynucleotide encoding a fusion protein described herein.

Table 17

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical	(*e.g.*, electroporation, particle gun,	YES	Transient	NO	Nucleic Acids and Proteins

Delivery	Vector/Mode	Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
	Calcium Phosphate transfection				
Viral	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modification	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	Cationic Liposomes	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral Delivery Vehicles	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids
	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

In another aspect, the delivery of genome editing system components or nucleic acids encoding such components, for example, a nucleic acid binding protein such as, for example, Cas9 or variants thereof, and a gRNA targeting a genomic nucleic acid sequence of interest, may be accomplished by delivering a ribonucleoprotein (RNP) to cells. The RNP comprises the nucleic acid binding protein, *e.g.*, Cas9, in complex with the targeting gRNA. RNPs may be delivered to cells using known methods, such as electroporation, nucleofection, or cationic lipid-mediated methods, for example, as reported by Zuris, J.A. *et al.*, 2015, *Nat. Biotechnology*, 33(1):73-80. RNPs are advantageous for use in CRISPR base editing

systems, particularly for cells that are difficult to transfect, such as primary cells. In addition, RNPs can also alleviate difficulties that may occur with protein expression in cells, especially when eukaryotic promoters, *e.g.*, CMV or EF1A, which may be used in CRISPR plasmids, are not well-expressed. Advantageously, the use of RNPs does not require the delivery of foreign DNA into cells. Moreover, because an RNP comprising a nucleic acid binding protein and gRNA complex is degraded over time, the use of RNPs has the potential to limit off-target effects. In a manner similar to that for plasmid based techniques, RNPs can be used to deliver binding protein (*e.g.*, Cas9 variants) and to direct homology directed repair (HDR).

A promoter used to drive base editor coding nucleic acid molecule expression can include AAV ITR. This can be advantageous for eliminating the need for an additional promoter element, which can take up space in the vector. The additional space freed up can be used to drive the expression of additional elements, such as a guide nucleic acid or a selectable marker. ITR activity is relatively weak, so it can be used to reduce potential toxicity due to over expression of the chosen nuclease.

Any suitable promoter can be used to drive expression of the base editor and, where appropriate, the guide nucleic acid. For ubiquitous expression, promoters that can be used include CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc. For brain or other CNS cell expression, suitable promoters can include: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc. For liver cell expression, suitable promoters include the Albumin promoter. For lung cell expression, suitable promoters can include SP-B. For endothelial cells, suitable promoters can include ICAM. For hematopoietic cells suitable promoters can include IFNbeta or CD45. For Osteoblasts suitable promoters can include OG-2.

In some embodiments, a base editor of the present disclosure is of small enough size to allow separate promoters to drive expression of the base editor and a compatible guide nucleic acid within the same nucleic acid molecule. For instance, a vector or viral vector can comprise a first promoter operably linked to a nucleic acid encoding the base editor and a second promoter operably linked to the guide nucleic acid.

The promoter used to drive expression of a guide nucleic acid can include: Pol III promoters such as U6 or H1 Use of Pol II promoter and intronic cassettes to express gRNA Adeno Associated Virus (AAV).

In particular embodiments, a fusion protein as described herein is encoded by a polynucleotide present in a viral vector (*e.g.*, adeno-associated virus (AAV), AAV3, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh8, AAV10, and variants thereof), or a suitable capsid protein of any viral vector. Thus, in some aspects, the disclosure relates to the viral delivery of a fusion protein. Examples of viral vectors include retroviral vectors (*e.g.* Maloney murine leukemia virus, MML-V), adenoviral vectors (*e.g.* AD100), lentiviral vectors (HIV and FIV-based vectors), herpesvirus vectors (*e.g.* HSV-2).

Viral Vectors

A base editor described herein can therefore be delivered with viral vectors. In some embodiments, a base editor disclosed herein can be encoded on a nucleic acid that is contained in a viral vector. In some embodiments, one or more components of the base editor system can be encoded on one or more viral vectors. For example, a base editor and guide nucleic acid can be encoded on a single viral vector. In other embodiments, the base editor and guide nucleic acid are encoded on different viral vectors. In either case, the base editor and guide nucleic acid can each be operably linked to a promoter and terminator. The combination of components encoded on a viral vector can be determined by the cargo size constraints of the chosen viral vector.

The use of RNA or DNA viral based systems for the delivery of a base editor takes advantage of highly evolved processes for targeting a virus to specific cells in culture or in the host and trafficking the viral payload to the nucleus or host cell genome. Viral vectors can be administered directly to cells in culture, patients (*in vivo*), or they can be used to treat cells *in vitro*, and the modified cells can optionally be administered to patients (*ex vivo*). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

Viral vectors can include lentivirus (*e.g.*, HIV and FIV-based vectors), Adenovirus (*e.g.*, AD100), Retrovirus (*e.g.*, Maloney murine leukemia virus, MML-V), herpesvirus vectors (*e.g.*, HSV-2), and Adeno-associated viruses (AAVs), or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No. 8,404,658 (formulations,

doses for AAV) and U.S. Patent No. 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For example, for AAV, the route of administration, formulation and dose can be as in U.S. Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Patent No. 5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (*e.g.* a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (*e.g.*, physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific base editing, the expression of the base editor and optional guide nucleic acid can be driven by a cell-type specific promoter.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (*See, e.g.*, Buchscher *et al.*, J. Virol. 66:2731-2739 (1992); Johann *et al.*, J. Virol. 66:1635-1640 (1992); Sommnerfelt *et al.*, Virol. 176:58-59 (1990); Wilson *et al.*, J. Virol. 63:2374-2378 (1989); Miller *et al.*, J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

Retroviral vectors, especially lentiviral vectors, can require polynucleotide sequences smaller than a given length for efficient integration into a target cell. For example, retroviral vectors of length greater than 9 kb can result in low viral titers compared with those of smaller size. In some aspects, a base editor of the present disclosure is of sufficient size so as

to enable efficient packaging and delivery into a target cell via a retroviral vector. In some embodiments, a base editor is of a size so as to allow efficient packing and delivery even when expressed together with a guide nucleic acid and/or other components of a targetable nuclease system.

Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and psi.2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, Adeno-associated virus ("AAV") vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA can be packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line can also be infected with adenovirus as a helper. The helper virus can promote replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid in some cases is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

In applications where transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors can also be used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (*See, e.g.*, West *et al.*, Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). The construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin *et al.*, Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, *et al.*, Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski *et al.*, J. Virol. 63:03822-3828 (1989).

AAV is a small, single-stranded DNA dependent virus belonging to the parvovirus family. The 4.7 kb wild-type (wt) AAV genome is made up of two genes that encode four replication proteins and three capsid proteins, respectively, and is flanked on either side by 145-bp inverted terminal repeats (ITRs). The virion is composed of three capsid proteins, Vp1, Vp2, and Vp3, produced in a 1:1:10 ratio from the same open reading frame but from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. A phospholipase domain, which functions in viral infectivity, has been identified in the unique N terminus of Vp1.

Similar to wt AAV, recombinant AAV (rAAV) utilizes the *cis*-acting 145-bp ITRs to flank vector transgene cassettes, providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can express a fusion protein as described herein and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. Although there are numerous examples of rAAV success using this system, *in vitro* and *in vivo*, the limited packaging capacity has limited the use of AAV-mediated gene delivery when the length of the coding sequence of the gene is equal or greater in size than the wt AAV genome.

Viral vectors can be selected based on the application. For example, for *in vivo* or *ex vivo* gene delivery, AAV can be advantageous over other viral vectors. In some embodiments, AAV allows low toxicity, which can be due to the purification method not requiring ultra-centrifugation of cell particles that can activate the immune response. In some embodiments, AAV allows low probability of causing insertional mutagenesis because it does not integrate into the host genome. Adenoviruses are commonly used as vaccines because of the strong immunogenic response they induce. Packaging capacity of the viral vectors can limit the size of the base editor that can be packaged into the vector.

AAV has a packaging capacity of about 4.5 Kb or 4.75 Kb including two 145 base inverted terminal repeats (ITRs). This means disclosed base editor as well as a promoter and transcription terminator can fit into a single viral vector. Constructs larger than 4.5 or 4.75 Kb can lead to significantly reduced virus production. For example, SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore, embodiments of the present disclosure include utilizing a disclosed base editor which is shorter in length than conventional base editors. In some examples, the base editors are less than 4 kb. Disclosed base editors can be less than 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb,

3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb. In some embodiments, the disclosed base editors are 4.5 kb or less in length.

An AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the type of AAV with regard to the cells to be targeted; *e.g.*, one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. A tabulation of certain AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)).

In some embodiments, lentiviral vectors are used. Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

Lentiviruses can be prepared as follows. After cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media is changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells are transfected with 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5 μg of psPAX2 (gag/pol/rev/tat). Transfection can be done in 4 mL OptiMEM with a cationic lipid delivery agent (50 μl Lipofectamine 2000 and 100 μl Plus reagent). After 6 hours, the media is changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

Lentivirus can be purified as follows. Viral supernatants are harvested after 48 hours. Supernatants are first cleared of debris and filtered through a 0.45 μm low protein binding (PVDF) filter. They are then spun in an ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets are resuspended in 50 μl of DMEM overnight at 4° C. They are then aliquoted and immediately frozen at -80°C.

In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated. In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses

angiostatic proteins endostatin and angiostatin that is contemplated to be delivered via a subretinal injection. In another embodiment, use of self-inactivating lentiviral vectors are contemplated.

Any RNA of the systems, for example a guide RNA or a base editor-encoding mRNA, can be delivered in the form of RNA. Base editor-encoding mRNA can be generated using *in vitro* transcription. For example, nuclease mRNA can be synthesized using a PCR cassette containing the following elements: T7 promoter, optional kozak sequence (GCCACC), nuclease sequence, and 3' UTR such as a 3' UTR from beta globin-polyA tail. The cassette can be used for transcription by T7 polymerase. Guide polynucleotides (*e.g.*, gRNA) can also be transcribed using *in vitro* transcription from a cassette containing a T7 promoter, followed by the sequence "GG", and guide polynucleotide sequence.

To enhance expression and reduce possible toxicity, the base editor-coding sequence and/or the guide nucleic acid can be modified to include one or more modified nucleoside *e.g.* using pseudo-uridine, 5-Methyl-cytosine, 2′-O-methyl-3′-phosphonoacetate, 2′-O-methyl ('M'), 2′-O-methyl-3′-phosphorothioate ('MS') and 2′-O-methyl-3′-thiophosphonoacetate ('MSP'), 5-methoxyuridine, phosphorothioate, and N1-Methylpseudouridine. Methods for using chemically modified mRNAs and guide RNAs are known in the art and described, for example, by Jiang et al., Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nat Commun 11, 1979 (2020). https://doi.org/10.1038/s41467-020-15892-8, Callum et al., *N*1-Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells, *Nucleic Acids Research*, Volume 48, Issue 6, 06 April 2020, Page e35, and Andries et al., Journal of Controlled Release, Volume 217, 10 November 2015, Pages 337-344, each of which is incorporated herein by reference in its entirety.

The small packaging capacity of AAV vectors makes the delivery of a number of genes that exceed this size and/or the use of large physiological regulatory elements challenging. These challenges can be addressed, for example, by dividing the protein(s) to be delivered into two or more fragments, wherein the N-terminal fragment is fused to a split intein-N and the C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or more AAV vectors. As used herein, "intein" refers to a self-splicing protein intron (*e.g.*, peptide) that ligates flanking N-terminal and C-terminal exteins (*e.g.*, fragments to be joined). The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014). For

example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

A fragment of a fusion protein as described herein can vary in length. In some embodiments, a protein fragment ranges from 2 amino acids to about 1000 amino acids in length. In some embodiments, a protein fragment ranges from about 5 amino acids to about 500 amino acids in length. In some embodiments, a protein fragment ranges from about 20 amino acids to about 200 amino acids in length. In some embodiments, a protein fragment ranges from about 10 amino acids to about 100 amino acids in length. Suitable protein fragments of other lengths will be apparent to a person of skill in the art.

In one embodiment, dual AAV vectors are generated by splitting a large transgene expression cassette in two separate halves (5′ and 3′ ends, or head and tail), where each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette is then achieved upon co-infection of the same cell by both dual AAV vectors followed by: (1) homologous recombination (HR) between 5′ and 3′ genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-head concatemerization of 5′ and 3′ genomes (dual AAV *trans*-splicing vectors); or (3) a combination of these two mechanisms (dual AAV hybrid vectors). The use of dual AAV vectors *in vivo* results in the expression of full-length proteins. The use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of >4.7 kb in size.

Inteins

Inteins (intervening protein) are auto-processing domains found in a variety of diverse organisms, which carry out a process known as protein splicing. Protein splicing is a multi-step biochemical reaction comprised of both the cleavage and formation of peptide bonds. While the endogenous substrates of protein splicing are proteins found in intein-containing organisms, inteins can also be used to chemically manipulate virtually any polypeptide backbone. Exemplary intein polypeptide sequences are provided in the Sequence Listing as SEQ ID NOs: 381-388.

In protein splicing, the intein excises itself out of a precursor polypeptide by cleaving two peptide bonds, thereby ligating the flanking extein (external protein) sequences via the

formation of a new peptide bond. This rearrangement occurs post-translationally (or possibly co-translationally). Intein-mediated protein splicing occurs spontaneously, requiring only the folding of the intein domain.

About 5% of inteins are split inteins, which are transcribed and translated as two separate polypeptides, the N-intein and C-intein, each fused to one extein. Upon translation, the intein fragments spontaneously and non-covalently assemble into the canonical intein structure to carry out protein splicing in trans. The mechanism of protein splicing entails a series of acyl-transfer reactions that result in the cleavage of two peptide bonds at the intein-extein junctions and the formation of a new peptide bond between the N- and C-exteins. This process is initiated by activation of the peptide bond joining the N-extein and the N-terminus of the intein. Virtually all inteins have a cysteine or serine at their N-terminus that attacks the carbonyl carbon of the C-terminal N-extein residue. This N to O/S acyl-shift is facilitated by a conserved threonine and histidine (referred to as the TXXH motif), along with a commonly found aspartate, which results in the formation of a linear (thio)ester intermediate. Next, this intermediate is subject to trans-(thio)esterification by nucleophilic attack of the first C-extein residue (+1), which is a cysteine, serine, or threonine. The resulting branched (thio)ester intermediate is resolved through a unique transformation: cyclization of the highly conserved C-terminal asparagine of the intein. This process is facilitated by the histidine (found in a highly conserved HNF motif) and the penultimate histidine and may also involve the aspartate. This succinimide formation reaction excises the intein from the reactive complex and leaves behind the exteins attached through a non-peptidic linkage. This structure rapidly rearranges into a stable peptide bond in an intein-independent fashion.

In some embodiments, a portion or fragment of a nuclease (*e.g.*, Cas9) is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (*e.g.*, nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, an N-terminal fragment of a base editor (*e.g.*, ABE, CBE) is fused to a split intein-N and a C-terminal fragment is fused to a split intein-C. These fragments are then packaged into two or more AAV vectors. In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

In one embodiment, inteins are utilized to join fragments or portions of an adenosine deaminase base editor protein that is grafted onto an AAV capsid protein. The use of certain inteins for joining heterologous protein fragments is described, for example, in Wood *et al.*, J. Biol. Chem. 289(21); 14512-9 (2014). For example, when fused to separate protein fragments, the inteins IntN and IntC recognize each other, splice themselves out and simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments. Other suitable inteins will be apparent to a person of skill in the art.

In some embodiments, an ABE was split into N- and C- terminal fragments at Ala, Ser, Thr, or Cys residues within selected regions of SpCas9. These regions correspond to loop regions identified by Cas9 crystal structure analysis. The N-terminus of each fragment is fused to an intein-N and the C- terminus of each fragment is fused to an intein C at amino acid positions S303, T310, T313, S355, A456, S460, A463, T466, S469, T472, T474, C574, S577, A589, and S590, which are indicated in bold capital letters in the sequence below.

```
  1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntdr hsikknliga llfdsgetae
 61 atrlkrtarr rytrrknric ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlnpdnsd
181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknglfgn
241 lialslgltp nfksnfdlae daklqlskdt ydddldnlla qigdqyadlf laaknlsdai
301 ll**S**dilrvn**T** ei**T**kaplsas mikrydehhq dltllkalvr qqlpekykei ffdq**S**kngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelh
421 ailrrqedfy pflkdnreki ekiltfripy yvgpl**A**rgn**S** rf**A**wm**T**rk**S**e e**T**i**T**pwnfee
481 vvdkgasaqs fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
541 sgeqkkaivd llfktnrkvt vkqlkedyfk kie**C**fd**S**vei sgvedrfn**AS** lgtyhdllki
601 ikdkdfldne enediledIv ltltlfedre mieerlktya hlfddkvmkq lkrrrytgwg
661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediqk aqvsgqgdsl
721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
841 ivpqsflkdd sidnkvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrmn tkydendkli revkvitlks
961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvrk
1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmp qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggfdsptva
1141 ysvlvvakve kgkskklksv kellgitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
1261 qhkhyldeii eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltnlga
```

1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqlggd (SEQ ID NO: 197).

USE OF NUCLEOBASE EDITORS

The correction of point mutations in disease-associated genes and alleles provides new strategies for gene correction with applications in therapeutics and basic research.

The present disclosure provides methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation that can be corrected by a base editor system provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, *e.g.*, a disease caused by a genetic mutation, an effective amount of a nucleobase editor (*e.g.*, an adenosine deaminase base editor) that corrects the point mutation in the disease associated gene. The present disclosure provides methods for the treatment of GSD1a that are associated with or caused by a point mutation that can be corrected by deaminase mediated gene editing. Suitable diseases that can be treated with the strategies and fusion proteins provided herein will be apparent to those of skill in the art based on the instant disclosure.

Provided herein are methods of using a base editor or base editor system for editing a nucleobase in a target nucleotide sequence associated with a disease or disorder (*e.g.*, GSD1a). In some embodiments, the activity of the base editor (*e.g.*, comprising an adenosine deaminase and a Cas9 domain) results in a correction of the point mutation. In some embodiments, the target DNA sequence comprises a G→A point mutation associated with a disease or disorder, and deamination of the mutant A base results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence comprises a T→C point mutation associated with a disease or disorder, and the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

In some embodiments, the deaminases (*e.g.*, adenosine deaminase) provided herein are capable of deaminating a deoxyadenosine residue of DNA. Other aspects of the disclosure provide fusion proteins that comprise a deaminase (*e.g.*, an adenosine deaminase) and a domain (*e.g.*, a Cas9) capable of binding to a specific nucleotide sequence. For example, adenosine can be converted to an inosine residue, which typically base pairs with a cytosine residue. Such fusion proteins are useful *inter alia* for targeted editing of nucleic acid sequences. Such fusion proteins can be used for targeted editing of DNA *in vitro*, *e.g.*, for the generation of mutant cells or animals; for the introduction of targeted mutations, *e.g.*, for the

correction of genetic defects in cells *ex vivo*, *e.g.*, in cells obtained from a subject that are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations *in vivo*, *e.g.*, the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a G to A, or a T to C to mutation can be treated using the nucleobase editors provided herein. The present disclosure provides fusion proteins, nucleic acids, vectors, compositions, methods, kits, systems, etc. that utilize the deaminases and nucleobase editors.

Use of Nucleobase Editors to Target Nucleotides in the G6PC gene

The suitability of nucleobase editors that target a nucleotide in the G6PC gene is evaluated as described herein. In one embodiment, a single cell of interest is transfected, transduced, or otherwise modified with a nucleic acid molecule or molecules encoding a nucleobase editor described herein together with a small amount of a vector encoding a reporter (*e.g.*, GFP). These cells can be immortalized human cell lines, such as 293T, K562 or U20S. Alternatively, primary human cells may be used. Cells may also be obtained from a subject or individual, such as from tissue biopsy, surgery, blood, plasma, serum, or other biological fluid. Such cells may be relevant to the eventual cell target,

Delivery may be performed using a viral vector as further described below. In one embodiment, transfection may be performed using lipid transfection (such as Lipofectamine, Metafectamine, or Fugene) or by electroporation. Following transfection, expression of GFP can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different nucleobase editors to determine which combinations of editors give the greatest activity.

The activity of the nucleobase editor is assessed as described herein, *i.e.*, by sequencing the target gene to detect alterations in the target sequence. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, miniprepped and sequenced with a single primer. Sequencing may also be performed using next generation sequencing techniques. When using next generation sequencing, amplicons may be 300-500 bp with the intended cut site placed asymmetrically. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, *e.g.*, for use in high throughput sequencing (for example on an Illumina MiSeq).

The fusion proteins that induce the greatest levels of target specific alterations in initial tests can be selected for further evaluation.

In particular embodiments, the nucleobase editors are used to target polynucleotides of interest. In one embodiment, a nucleobase editor as described herein is delivered to cells (*e.g.*, hepatocytes) in conjunction with a guide RNA that is used to target a nucleic acid sequence, *e.g.*, a G6PC polynucleotide harboring GSD1a-associated mutations, thereby altering the target gene, *i.e.*, G6PC.

In some embodiments, a base editor is targeted by a guide RNA to introduce one or more edits to the sequence of a gene of interest (*e.g.* G6PC). In some embodiments, the one or more alterations are introduced into the glucose-6-phosphatase (G6PC) gene. In some embodiments the one or more alterations is R83C. In some embodiments, the one or more alterations is Q347X. In some embodiments, the alteration is introduced into a representative *Homo sapiens* G6PC protein, found under NCBI Reference Sequence No. AAA16222.1. In some embodiments, the alteration is introduced into a representative *Homo sapiens* G6PC nucleic acid sequence, found under GenBank Reference Sequence No. U01120.1.

Methods for Editing Nucleic Acids

Some aspects of the disclosure provide methods for editing a nucleic acid. In some embodiments, the method is a method for editing a nucleobase of a nucleic acid (*e.g.*, a base pair of a double-stranded DNA sequence). In some embodiments, the method comprises the steps of: a) contacting a target region of a nucleic acid (*e.g.*, a double-stranded DNA sequence) with a complex comprising a base editor (*e.g.*, a Cas9 domain fused to an adenosine deaminase) and a guide nucleic acid (*e.g.*, gRNA), wherein the target region comprises a targeted nucleobase pair, b) inducing strand separation of said target region, c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase, and d) cutting no more than one strand of said target region, where a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase. In some embodiments, the method results in less than 20% indel formation in the nucleic acid. It should be appreciated that in some embodiments, step b is omitted. In some embodiments, the method results in less than 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2%, or less than 0.1% indel formation. In some embodiments, the method further comprises replacing the second nucleobase with a fifth nucleobase that is complementary to the fourth nucleobase, thereby

generating an intended edited base pair (*e.g.*, C•G to T•A). In some embodiments, at least 5% of the intended base pairs are edited. In some embodiments, at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, or 50% of the intended base pairs are edited.

In some embodiments, the ratio of intended products to unintended products in the target nucleotide is at least 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, or 200:1, or more. In some embodiments, the ratio of intended mutation to indel formation is greater than 1:1, 10:1, 50:1, 100:1, 500:1, or 1000:1, or more. In some embodiments, the cut single strand (nicked strand) is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the strand comprising the first nucleobase. In some embodiments, the base editor comprises a Cas9 domain. In some embodiments, the base editor protects or binds the non-edited strand. In some embodiments, the base editor comprises nickase activity. In some embodiments, the intended edited base pair is upstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edited base pair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site. In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker. In some embodiments, the linker is 1-25 amino acids in length. In some embodiments, the linker is 5-20 amino acids in length. In some embodiments, linker is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length. In one embodiment, the linker is 32 amino acids in length. In another embodiment, a "long linker" is at least about 60 amino acids in length. In other embodiments, the linker is between about 3-100 amino acids in length. In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1-10 nucleotides. In some embodiments, the target window is 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, or 1 nucleotides in length. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edited base pair is within the target window. In some embodiments, the target window comprises the intended edited base pair. In some embodiments, the method is performed using any of the base editors provided herein. In some embodiments, a target window is a methylation window.

In some embodiments, the disclosure provides methods for editing a nucleotide. In some embodiments, the disclosure provides a method for editing a nucleobase pair of a double-stranded DNA sequence. In some embodiments, the method comprises a) contacting a target region of the double-stranded DNA sequence with a complex comprising a base editor and a guide nucleic acid (*e.g.*, gRNA), where the target region comprises a target nucleobase pair, b) inducing strand separation of said target region, c) converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase, d) cutting no more than one strand of said target region, wherein a third nucleobase complementary to the first nucleobase base is replaced by a fourth nucleobase complementary to the second nucleobase, and the second nucleobase is replaced with a fifth nucleobase that is complementary to the fourth nucleobase, thereby generating an intended edited base pair, wherein the efficiency of generating the intended edited base pair is at least 5%. It should be appreciated that in some embodiments, step b is omitted. In some embodiments, at least 5% of the intended base pairs are edited. In some embodiments, at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, or 50% of the intended base pairs are edited. In some embodiments base editing by a method described herein may have a base conversion efficiency of at least 10% at any particular gene site. In some embodiments, base editing by a method described herein may have a base conversion efficiency of at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50% at least 55% or at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95%, 96%, 97%, 98% or at least 99% at any particular gene site. In some embodiments base editing by a method described herein may have a base conversion efficiency of at least 70% at any particular gene site. In some embodiments base editing by a method described herein may have a base conversion efficiency of at least 80% at any particular gene site. In some embodiments base editing by a method described herein may have a base conversion efficiency of at least 90% at any particular gene site.

In some embodiments, the method causes less than 19%, 18%, 16%, 14%, 12%, 10%, 8%, 6%, 4%, 2%, 1%, 0.5%, 0.2%, or less than 0.1% indel formation. In some embodiments, the ratio of intended product to unintended products at the target nucleotide is at least 2:1, 5:1, 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, or 200:1, or more. In some embodiments, the ratio of intended mutation to indel formation is greater than 1:1, 10:1, 50:1, 100:1, 500:1, or 1000:1, or more. In some embodiments, the cut single strand is hybridized to the guide nucleic acid. In some embodiments, the cut single strand is opposite to the

strand comprising the first nucleobase. In some embodiments, the nucleobase editor comprises nickase activity. In some embodiments, the intended edited base pair is upstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. In some embodiments, the intended edited base pair is downstream of a PAM site. In some embodiments, the intended edited base pair is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream stream of the PAM site. In some embodiments, the method does not require a canonical (*e.g.*, NGG) PAM site. In some embodiments, the nucleobase editor comprises a linker. In some embodiments, the linker is 1-25 amino acids in length. In some embodiments, the linker is 5-20 amino acids in length. In some embodiments, the linker is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length. *e.g.*, In some embodiments, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. In some embodiments, the target window comprises 1-10 nucleotides. In some embodiments, the target window is 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, or 1 nucleotides in length. In some embodiments, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some embodiments, the intended edited base pair occurs within the target window. In some embodiments, the target window comprises the intended edited base pair. In some embodiments, the nucleobase editor is any one of the base editors provided herein.

Expression of Fusion Proteins in a Host Cell

Fusion proteins of the disclosure comprising an adenosine deaminase variant may be expressed in virtually any host cell of interest, including but not limited to, bacteria, yeast, fungi, insects, plants, and animal cells, using routine methods known to the skilled artisan. For example, a DNA encoding an adenosine deaminase of the disclosure can be cloned by designing suitable primers for the upstream and downstream of CDS based on the cDNA sequence. The cloned DNA may be directly, or after digestion with a restriction enzyme when desired, or after addition of a suitable linker and/or a nuclear localization signal ligated with a DNA encoding one or more additional components of a base editing system. The base editing system is translated in a host cell to form a complex.

A DNA encoding a protein domain described herein can be obtained by chemically synthesizing the DNA, or by connecting synthesized partly overlapping oligoDNA short chains by utilizing the PCR method and the Gibson Assembly method to construct a DNA

encoding the full length thereof. The advantage of constructing a full-length DNA by chemical synthesis or a combination of PCR method or Gibson Assembly method is that the codon to be used can be designed in CDS full-length according to the host into which the DNA is introduced. In the expression of a heterologous DNA, the protein expression level is expected to increase by converting the DNA sequence thereof to a codon highly frequently used in the host organism. As the data of codon use frequency in host to be used, for example, the genetic code use frequency database (kazusa.or.jp/codon/index.html) disclosed in the home page of Kazusa DNA Research Institute can be used, or documents showing the codon use frequency in each host may be referred to. By reference to the obtained data and the DNA sequence to be introduced, codons showing low use frequency in the host from among those used for the DNA sequence may be converted to a codon coding the same amino acid and showing high use frequency.

An expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme can be produced, for example, by linking the DNA to the downstream of a promoter in a suitable expression vector. In some embodiments, animal cell expression plasmids (*e.g.*, pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNAI/Neo), and animal virus vectors such as retrovirus, vaccinia virus, adenovirus, and the like are used. In other embodiments, *Escherichia coli*-derived plasmids (*e.g.*, pBR322, pBR325, pUC12, pUC13); *Bacillus subtilis*-derived plasmids (*e.g.*, pUB110, pTP5, pC194); yeast-derived plasmids (*e.g.*, pSH19, pSH15); insect cell expression plasmids (*e.g.*, pFast-Bac); bacteriophages such as lambda phage and the like; insect virus vectors such as baculovirus and the like (*e.g.*, BmNPV, AcNPV); and the like are used.

In some embodiments, any promoter appropriate for gene expression in a given host can be used. In a conventional method using DSB, since the survival rate of the host cell sometimes decreases markedly due to the toxicity, it is desirable to increase the number of cells by the start of the induction by using an inductive promoter. However, since sufficient cell proliferation can also be afforded by expressing the nucleic acid-modifying enzyme complex of the present disclosure, a constitution promoter can also be used without limitation.

For example, without limitation, when the host is an animal cell, SRα promoter, SV40 promoter, LTR promoter, CMV (cytomegalovirus) promoter, RSV (*Rous sarcoma* virus) promoter, MoMuLV (*Moloney* mouse leukemia virus) LTR, HSV-TK (simple herpes virus

thymidine kinase) promoter and the like are used. Of these, CMV promoter, SRα promoter and the like are suitable for use.

In addition to those mentioned above, an expression vector containing an enhancer, splicing signal, terminator, polyA addition signal, a selection marker such as drug resistance gene, auxotrophic complementary gene and the like, replication origin and the like, on demand can be used.

An RNA encoding a protein domain described herein can be prepared, for example, by transcription to mRNA in an *in vitro* transcription system known *per se* by using a vector encoding DNA encoding the above-mentioned nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme as a template.

A fusion protein of the disclosure can be intracellularly expressed by introducing an expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme into a host cell, and culturing the host cell.

As the animal cell, cell lines such as monkey COS-7 cell, monkey Vero cell, Chinese hamster ovary (CHO) cell, dhfr gene-deficient CHO cell, mouse L cell, mouse AtT-20 cell, mouse myeloma cell, rat GH3 cell, human FL cell and the like, pluripotent stem cells such as iPS cell, ES cell and the like of human and other mammals, and primary cultured cells prepared from various tissues are used. Furthermore, zebrafish embryo, *Xenopus* oocyte and the like can also be used.

All the above-mentioned host cells may be haploid (monoploid), or polyploid (*e.g.*, diploid, triploid, tetraploid and the like). In the conventional mutation introduction methods, mutation is, in principle, introduced into only one homologous chromosome to produce a hetero gene type. Therefore, desired phenotype is not expressed unless dominant mutation occurs, and homozygousness inconveniently requires labor and time. In contrast, according to the present disclosure, since mutation can be introduced into any allele on the homologous chromosome in the genome, the desired phenotype can be expressed in a single generation even in the case of recessive mutation, which is extremely useful since the problem of the conventional method can be solved.

An expression vector can be introduced by a known method (*e.g.*, lysozyme method, competent method, PEG method, $CaCl_2$ coprecipitation method, electroporation method, the microinjection method, the particle gun method, lipofection method, Agrobacterium method and the like) according to the kind of the host.

A vector can be introduced into an animal cell according to the methods described in, for example, *Cell Engineering* additional volume 8, New Cell Engineering Experiment Protocol, 263-267 (1995) (published by Shujunsha), and *Virology*, 52, 456 (1973). A cell comprising a vector can be cultured according to a known method according to the type of host.

As a medium for culturing an animal cell, for example, minimum essential medium (MEM) containing about 5 to about 20% of fetal bovine serum (*Science*, 122, 501 (1952)), Dulbecco's modified Eagle medium (DMEM) (*Virology*, 8, 396 (1959)), RPMI 1640 medium (*The Journal of the American Medical Association*, 199, 519 (1967)), 199 medium (*Proceeding of the Society for the Biological Medicine*, 73, 1 (1950)) and the like are used. The pH of the medium is preferably about 6 to about 8. The culture is generally maintained at about 30°C to about 40°C. Where necessary, aeration and stirring may be performed.

When a higher eukaryotic cell, such as animal cell, is used as a host cell, a DNA encoding a base editing system of the present disclosure (*e.g.*, comprising an adenosine deaminase variant) is introduced into a host cell under the regulation of an inducible promoter (*e.g.*, metallothionein promoter (induced by heavy metal ion), heat shock protein promoter (induced by heat shock), Tet-ON/Tet-OFF system promoter (induced by addition or removal of tetracycline or a derivative thereof), steroid-responsive promoter (induced by steroid hormone or a derivative thereof) etc.), the induction substance is added to the medium (or removed from the medium) at an appropriate stage to induce expression of the nucleic acid-modifying enzyme complex, culture is performed for a given period to carry out a base editing and, introduction of a mutation into a target gene, transient expression of the base editing system can be realized.

Alternatively, the above-mentioned inductive promoter can also be utilized as a vector removal mechanism when higher eukaryotic cells, such animal cells and the like, are used as host cells. That is, a vector is mounted with a replication origin that functions in a host cell, and a nucleic acid encoding a protein necessary for replication (*e.g.*, SV40 on and large T antigen, oriP and EBNA-1 etc. for animal cells) of the expression of the nucleic acid encoding the protein is regulated by the above-mentioned inducible promoter. As a result, while the vector is autonomously replicatable in the presence of an induction substance, when the induction substance is removed, autonomous replication is not available, and the vector naturally falls off along with cell division (autonomous replication is not possible by the addition of tetracycline and doxycycline in Tet-OFF system vector).

Precise Correction of Pathogenic Mutations

In some embodiments, the intended mutation is a precise correction of a pathogenic mutation or a disease-causing mutation. The pathogenic mutation can be a pathogenic single nucleotide polymorphism (SNP) or be caused by a SNP. For example, the pathogenic mutation can be an amino acid change in a protein encoded by a gene. In another example, the pathogenic mutation can be a pathogenic SNP in a gene. The precise correction can be to revert the pathogenic mutation back to its wild-type state. In some embodiments, the pathogenic mutation is a G→A point mutation associated with a disease or disorder, and wherein the deamination of the mutant A base with an A-to-G base editor (ABE) results in a sequence that is not associated with a disease or disorder. In some embodiments, the pathogenic mutation is a C→T point mutation. The C→T point mutation can be corrected, for example, by targeting an A-to-G base editor (ABE) to the opposite strand and editing the complement A of the pathogenic T nucleobase. A base editor can be targeted to a pathogenic SNP, or to the complement of the pathogenic SNP. The nomenclature of the description of pathogenic or disease-causing mutations and other sequence variations are described in den Dunnen, J.T. and Antonarakis, S.E., "Mutation Nomenclature Extensions and Suggestions to Describe Complex Mutations: A Discussion." Human Mutation 15:712 (2000), the entire contents of which is hereby incorporated by reference.

In a particular embodiment, the disease or disorder is Glycogen Storage Disease Type 1 (GSD1 or Von Gierke Disease). In some embodiments, the disease or disorder is GSD1a. In some embodiments, the pathogenic mutation is in the G6PC gene. In some embodiments, the pathogenic mutation is Q347X of the G6PC gene. In some embodiments, the pathogenic mutation is R83C of the G6PC gene.

Pharmaceutical Compositions

In some aspects, a pharmaceutical composition is provided, which comprises any of the base editors, fusion proteins, or the fusion protein-guide polynucleotide complexes described herein.

The pharmaceutical compositions as described herein can be prepared in accordance with known techniques. *See*, *e.g.*, Remington, The Science And Practice of Pharmacy (21st ed. 2005). In some embodiments, the pharmaceutical composition comprises a pharmaceutically acceptable carrier. Suitable pharmaceutically acceptable carriers generally comprise inert substances that aid in administering the pharmaceutical composition to a

subject, aid in processing the pharmaceutical compositions into deliverable preparations, or aid in storing the pharmaceutical composition prior to administration. Pharmaceutically acceptable carriers can include agents that can stabilize, optimize or otherwise alter the form, consistency, viscosity, pH, pharmacokinetics, solubility of the formulation. Such agents include buffering agents, wetting agents, emulsifying agents, diluents, encapsulating agents, and skin penetration enhancers. For example, carriers can include, but are not limited to, saline, buffered saline, dextrose, arginine, sucrose, water, glycerol, ethanol, sorbitol, dextran, sodium carboxymethyl cellulose, and combinations thereof.

Some nonlimiting examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents, such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum alcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the formulation.

Pharmaceutical compositions can comprise one or more pH buffering compounds to maintain the pH of the formulation at a predetermined level that reflects physiological pH, such as in the range of about 5.0 to about 8.0. The pH buffering compound used in the aqueous liquid formulation can be an amino acid or mixture of amino acids, such as histidine or a mixture of amino acids such as histidine and glycine. Alternatively, the pH buffering compound is preferably an agent which maintains the pH of the formulation at a predetermined level, such as in the range of about 5.0 to about 8.0, and which does not chelate calcium ions. Illustrative examples of such pH buffering compounds include, but are

not limited to, imidazole and acetate ions. The pH buffering compound may be present in any amount suitable to maintain the pH of the formulation at a predetermined level.

Pharmaceutical compositions can also contain one or more osmotic modulating agents, *i.e.*, a compound that modulates the osmotic properties (*e.g.*, tonicity, osmolality, and/or osmotic pressure) of the formulation to a level that is acceptable to the blood stream and blood cells of recipient individuals. The osmotic modulating agent can be an agent that does not chelate calcium ions. The osmotic modulating agent can be any compound known or available to those skilled in the art that modulates the osmotic properties of the formulation. One skilled in the art may empirically determine the suitability of a given osmotic modulating agent for use in the inventive formulation. Illustrative examples of suitable types of osmotic modulating agents include, but are not limited to: salts, such as sodium chloride and sodium acetate; sugars, such as sucrose, dextrose, and mannitol; amino acids, such as glycine; and mixtures of one or more of these agents and/or types of agents. The osmotic modulating agent(s) may be present in any concentration sufficient to modulate the osmotic properties of the formulation.

In some embodiments, the pharmaceutical composition is formulated for delivery to a subject, *e.g.*, for gene editing. Suitable routes of administrating the pharmaceutical composition described herein include, without limitation: topical, subcutaneous, suboccipital, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

In some embodiments, the pharmaceutical composition described herein is administered locally to a diseased site. In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, by means of a catheter, by means of a suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber.

In other embodiments, the pharmaceutical composition described herein is delivered in a controlled release system. In one embodiment, a pump can be used (see, *e.g.*, Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald *et al.*, 1980, Surgery 88:507; Saudek *et al.*, 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used. (See, *e.g.*, Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug

Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. See also Levy *et al.*, 1985, Science 228: 190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71: 105.) Other controlled release systems are discussed, for example, in Langer, supra.

In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a subject, *e.g.*, a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic use as solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

A pharmaceutical composition for systemic administration can be a liquid, *e.g.*, sterile saline, lactated Ringer's or Hank's solution. In addition, the pharmaceutical composition can be in solid forms and re-dissolved or suspended immediately prior to use. Lyophilized forms are also contemplated. The pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is also suitable for parenteral administration. The particles can be of any suitable structure, such as unilamellar or plurilamellar, so long as compositions are contained therein. Compounds can be entrapped in "stabilized plasmid-lipid particles" (SPLP) containing the fusogenic lipid dioleoylphosphatidylethanolamine (DOPE), low levels (5-10 mol%) of cationic lipid, and stabilized by a polyethyleneglycol (PEG) coating (Zhang Y. P. *et al.*, Gene Ther. 1999, 6: 1438-47). Positively charged lipids such as N-[l-(2,3-dioleoyloxi)propyl]-N,N,N-trimethyl-amoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. See, *e.g.*, U.S. Patent Nos. 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

The pharmaceutical composition described herein can be administered or packaged as a unit dose, for example. The term "unit dose" when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; *i.e.*, carrier, or vehicle.

Further, the pharmaceutical composition can be provided as a pharmaceutical kit comprising (a) a container containing a compound as described herein in lyophilized form and (b) a second container containing a pharmaceutically acceptable diluent (*e.g.*, sterile used for reconstitution or dilution of the lyophilized compound described herein. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use, or sale for human administration.

In another aspect, an article of manufacture containing materials useful for the treatment of the diseases described above is included. In some embodiments, the article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers can be formed from a variety of materials such as glass or plastic. In some embodiments, the container holds a composition that is effective for treating a disease described herein and can have a sterile access port. For example, the container can be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle. The active agent in the composition is a compound as described and provided herein. In some embodiments, the label on or associated with the container indicates that the composition is used for treating the disease of choice. The article of manufacture can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

In some embodiments, any of the fusion proteins, gRNAs, and/or complexes described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises any of the fusion proteins provided herein. In some embodiments, the pharmaceutical composition comprises any of the complexes provided

herein. In some embodiments, the pharmaceutical composition comprises a ribonucleoprotein complex comprising an RNA-guided nuclease (*e.g.*, Cas9) that forms a complex with a gRNA and a cationic lipid. In some embodiments pharmaceutical composition comprises a gRNA, a nucleic acid programmable DNA binding protein, a cationic lipid, and a pharmaceutically acceptable excipient. Pharmaceutical compositions can optionally comprise one or more additional therapeutically active substances.

In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and contacted with any of the pharmaceutical compositions provided herein. In some embodiments, cells removed from a subject and contacted *ex vivo* with a pharmaceutical composition are re-introduced into the subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering pharmaceutical compositions comprising nucleases are known, and are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, domesticated animals, pets, and commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

Formulations of the pharmaceutical compositions described herein can be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient(s) into association with

an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit. Pharmaceutical formulations can additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington's The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated in its entirety herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011/053982 A8, filed Nov. 2, 2010), incorporated in its entirety herein by reference, for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease.

Except insofar as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this disclosure.

The compositions, as described above, can be administered in effective amounts. The effective amount will depend upon the mode of administration, the particular condition being treated, and the desired outcome. It may also depend upon the stage of the condition, the age and physical condition of the subject, the nature of concurrent therapy, if any, and like factors well-known to the medical practitioner. For therapeutic applications, it is that amount sufficient to achieve a medically desirable result.

In some embodiments, compositions in accordance with the present disclosure can be used for treatment of any of a variety of diseases, disorders, and/or conditions.

Methods of Treating Glycogen Storage Disease Type 1a (GSD1a)

Provided also are methods of treating Glycogen Storage Disease Type 1a (GSD1a) and/or the genetic mutations in G6PC that cause GSD1a that comprise administering to a subject (*e.g.*, a mammal, such as a human) a therapeutically effective amount of a pharmaceutical composition that comprises a polynucleotide encoding a base editor system (*e.g.*, Adenosine Deaminase Base Editor (ABE) and gRNA) described herein. In some

embodiments, the base editor is a fusion protein that comprises a polynucleotide programmable DNA binding domain and an adenosine deaminase domain. A cell of the subject is transduced with the base editor and one or more guide polynucleotides that target the base editor to effect an A•T to G•C alteration of a nucleic acid sequence containing mutations in the G6PC gene.

The methods herein include administering to the subject (including a subject identified as being in need of such treatment, or a subject suspected of being at risk of disease and in need of such treatment) an effective amount of a composition described herein. Identifying a subject in need of such treatment can be in the judgment of a subject or a health care professional and can be subjective (*e.g.*, opinion) or objective (*e.g.*, measurable by a test or diagnostic method).

The therapeutic methods, in general, comprise administration of a therapeutically effective amount of a pharmaceutical composition comprising, for example, a vector encoding a base editor and a gRNA that targets the G6PC gene of a subject (*e.g.*, a human patient) in need thereof. Such treatment will be suitably administered to a subject, particularly a human subject, suffering from, having, susceptible to, or at risk for GSD1a. The compositions herein may be also used in the treatment of any other disorders in which GSD1a may be implicated.

In one embodiment, a method of monitoring treatment progress is provided. The method includes the step of determining a level of diagnostic marker (Marker) (*e.g.*, SNP associated with GSD1a) or diagnostic measurement (*e.g.*, screen, assay) in a subject suffering from or susceptible to a disorder or symptoms thereof associated with GSD1a in which the subject has been administered a therapeutic amount of a composition herein sufficient to treat the disease or symptoms thereof. The level of Marker determined in the method can be compared to known levels of Marker in either healthy normal controls or in other afflicted patients to establish the subject's disease status. In preferred embodiments, a second level of Marker in the subject is determined at a time point later than the determination of the first level, and the two levels are compared to monitor the course of disease or the efficacy of the therapy. In certain preferred embodiments, a pre-treatment level of Marker in the subject is determined prior to beginning treatment or therapy as described herein; this pre-treatment level of Marker can then be compared to the level of Marker in the subject after the treatment or therapy commences, to determine the efficacy of the treatment or therapy.

In some embodiments, cells are obtained from the subject and contacted with a pharmaceutical composition as provided herein. In some embodiments, cells removed from a subject and contacted *ex vivo* with a pharmaceutical composition are re-introduced into the subject, optionally after the desired genomic modification has been affected or detected in the cells.

Methods of delivering pharmaceutical compositions comprising nucleases are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties. Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts, for example, for veterinary use.

Kits

Various aspects of this disclosure provide kits comprising a base editor system. In one embodiment, the kit comprises a nucleic acid construct comprising a nucleotide sequence encoding a nucleobase editor fusion protein. The fusion protein comprises a deaminase (*e.g.*, adenine deaminase) and a nucleic acid programmable DNA binding protein (napDNAbp). In some embodiments, the kit comprises at least one guide RNA capable of targeting a nucleic acid molecule of interest, *e.g.*, G6PC GSD1a associated mutations. In some embodiments, the kit comprises a nucleic acid construct comprising a nucleotide sequence encoding at least one guide RNA.

The kit provides, in some embodiments, instructions for using the kit to edit one or more G6PC GSD1a associated mutations**.** The instructions will generally include information about the use of the kit for editing nucleic acid molecules. In other embodiments, the instructions include at least one of the following: precautions; warnings; clinical studies; and/or references. The instructions may be printed directly on the container (when present), or as a label applied to the container, or as a separate sheet, pamphlet, card, or folder supplied in or with the container. In a further embodiment, a kit can comprise instructions in the form of a label or separate insert (package insert) for suitable operational parameters. In yet another embodiment, the kit can comprise one or more containers with appropriate positive and negative controls or control samples, to be used as standard(s) for

detection, calibration, or normalization. The kit can further comprise a second container comprising a pharmaceutically-acceptable buffer, such as (sterile) phosphate-buffered saline, Ringer's solution, or dextrose solution. It can further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use. In certain embodiments, the kit is useful for the treatment of a subject having Glycogen Storage Disease Type 1a (GSD1a).

The practice of the embodiments described and provided herein employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides described and provided herein, and, as such, may be considered in making and practicing the disclosed and described embodiments. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods described herein, and are not intended to limit the scope of any of the embodiments and/or the disclosure described herein.

EXAMPLES

Example 1: Precise correction in vivo in heterozygous transgenic Glycogen storage disease type 1a (GSD1a) R83C mice

Glycogen storage disease type 1a (GSD1a) is caused by a mutation in the glucose-6-phosphatase (G6PC) gene, which affects about 80% of patients with GSD1a. The R83C mutation affects about 900 US patients annually diagnosed with Glycogen storage disease type 1a (GSD1a). This mutation is a single base substitution that introduces a cysteine at

position 83 (R83C) of the G6PC protein. A precise correction of R83C will likely restore expression of G6PC and normalize glucose metabolism. A representative G6PC nucleotide target sequence (ATTCTCTTTGGACAGTGTCCATACTGGTGG (SEQ ID NO 399) having complementary sequence TAAGAGAAACCTGTCACAGGTATGACCACC (SEQ ID NO: 400): and corresponding amino acid sequence (ILFGQCPYWW (SEQ ID NO: 401))indicating on target and bystander site "a" nucleobases for correction of the R83C mutation are shown in **FIG. 1**. A precise correction at this site would yield the following conversion: TGT > CGT or TGT > CGC (Cysteine > Arginine).

The G6PC gRNA sequence hybridizes to the complement of the G6PC target sequence shown below:

CAGTATGGACACTGTCCAAAGAGAAT (SEQ ID NO: 395)

The NNGRRT PAM sequence (*i.e.*, *Staphylococcus aureus* Cas9 (saCas9)) is underlined above.

The gRNA sequence is as follows: CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370).

The base-editing efficiency of adenosine deaminase base editors (ABE) using TadA variants MSP605, MSP824, MSP825, MSP680, MSP828, and MSP829 (see **Table 18**) and saCas9n was evaluated *in vivo* using a transgenic mouse model heterozygous for huG6PC, harboring the R83C mutation for Glycogen storage disease type 1a (GSD1a) (**FIGs. 2B and 3**). The use of saCas9 for efficient *in vivo* genome editing and exemplification of an saCas9 sgRNA scaffold are described in A. Ran et al. (2015, *Nature*, Vol. 520, pages 186–191). The engineering of Cas9 variants, e.g., SaCas9, with relaxed PAM recognition specificities is described by B.P. Kleinstiver et al., 2014, *Nature Biotechnol.*, 33(12):1293-1299.

Table 18. Adenosine Deaminase Base Editor Variants

TadA Variant	mRNA base-editor variant
MSP605	dimeric TadA-ABE7.10 (Y147T+Q154S+V82G)-saCas9n
MSP824	dimeric TadA-ABE7.10 (Y147D+Q154S+V82G+F149Y+D167N)-saCas9n
MSP825	dimeric TadA-ABE7.10 (Y147D+Q154S+V82G+L36H+N157K+F149Y+D167N)-saCas9n
MSP680	monomeric TadA-ABE7.10 (Y147T+Q154S+V82G+I76Y)-saCas9n
MSP828	monomeric TadA-ABE7.10 (Y147D+Q154S+V82G+I76Y+F149Y+D167N)-saCas9n
MSP829	monomeric TadA-ABE7.10 (Y147D+Q154S+V82G+I76Y+L36H+N157K+F149Y+D167N)-saCas9n

FIG. 2A depicts the *in vivo* workflow used to introduce the base editors into the transgenic mice. Lipid nanoparticles (LNP) carrying base editor mRNA and gRNA were dosed via intravenous (IV) injection into the transgenic mice at a dose of 1 mg/kg. Next-generation sequencing data from whole-liver extracts revealed significant correction for R83C (**FIGs. 2B and 3**). TadA variant MSP828 demonstrated about 40% precise correction of the R83C mutation, with low bystander editing. This level of mutation correction is expected to restore glucose homeostasis.

The level of *in vivo* correction for Glycogen storage disease type 1a (GSD1a) by base-editing was greater than that achieved by HDR-base (homology directed repair) methodologies and was achieved without insertion or double-stranded breaks. These results demonstrate base-editing technology as an effective approach for mutation correction for Glycogen storage disease type 1a (GSD1a) and its therapeutic potential.

Example 2: In vivo base editing correction of metabolic defects in GSD1a R83C mice

GSD1a overview

As depticted schematically in FIG. 4, (GSD1a) is an autosomal recessive disorder caused by mutations in the *G6PC* gene. The most prevalent pathogenic mutation identified in Caucasian GSD1a patients is R83C, located in the active site of the enzyme and associated with inactivation of G6Pase. A loss of G6Pase function can result in life-threatening hypoglycemia, seizures and even death. To mitigate hypoglycemia, patients must maintain strict and frequent adherence to glucose supplementation through day and night, by way of a slow glucose release formula. One missed or delayed dose can result in emergency hypoglycemia. Among many complications, enlarged liver, accumulation of uric acid, lactate, and lipids are common in GSD1a patients.

Utility of the described base editors for generating permanent and predictable single nucleotide substitutions

The R83C mutation introduces a single G>A conversion in the *g6pc* gene. Adenine base editors (ABEs) as described herein effect the programmable conversion of A to G in genomic DNA, thus supporting their utility to correct this mutation. As shown schematically in FIG. 5, the adenine base editor is a fusion protein containing an evolved TadA deaminase connected to CRISPR-Cas enzyme. The base editor binds to target DNA that is complementary to the guide-RNA (superimposed on the CRISPR-Cas9 enzyme) and exposes

a stretch of single-stranded DNA. The deaminase converts the target adenine into inosine, and the Cas enzyme nicks the opposite strand, which is then repaired, completing the base pair conversion. Thus, the direct repair of a point mutation has the potential for restoration of gene function.

In this Example, base-editors for A>G conversion in the *g6pc* gene were optimized for correction of R83C. Shown in FIG. 6A is the target DNA sequence (CCACCAGTATGGACACTGTCCAAAGAGAAT (SEQ ID NO: 402)) and underlying amino acid translation for the GSD1a R83C mutation (WWYPCQGFLI; SEQ ID NO: 403). The target nucleobase to be edited is represented by double underlining, at position 12. The editing window also includes a possible bystander, shown represented by single underlining at position 6. An edit that may result in a synonymous conversion is shown at position 10.

For screening, a HEK293 cell line that expressed the G6PC transgene harboring the R83C mutation was generated and was transfected with base-editor mRNA and gRNA. Allele frequencies were assessed by high-throughput targeted amplicon Next-Generation Sequencing. Variants 1-5 represent a combination of gRNA and base-editor RNA, engineered for optimized target correction. Variant 5 yielded approximately 60% targeted base-editing efficiency for R83C correction and limited bystander editing (FIG. 6B).

Mouse *in vivo* disease model and demonstration of *in vivo* correction of the R83C single nucleotide mutation

In vivo correction of R83C base editing

To validate base-editing efficiency for R83C correction *in vivo*, a novel GSD1a mouse that expresses the human G6PC-R83C transgene in place of mouse G6pc was generated. It was confirmed that mice homozygous for huR83C exhibited postnatal lethality and rarely survived to weaning (21 days). On glucose supplementation therapy, the animals survived to at least 3 weeks of age and revealed characteristic pathological signatures of GSD1a, such as reduced body weight, enlarged livers, significant G6Pase inhibition, and abnormal serum metabolites compared to littermate controls (FIG. 7). This phenotype is consistent with published and clinical reports in humans.

For the *in vivo* experiments, LNP-mediated delivery was tested in transgenic mice that were heterozygous for huR83C due to neonatal lethality of homozygous mice. The schematic in FIG. 2A depicts *in vivo* workflow, with lipid nanoparticle, or LNP, co-formulations of base-editor mRNA and gRNA dosed via IV injection. Given neonatal

lethality of the homozygous mice, LNP-dosing was administered via the temporal vein shortly post birth, and activity was compared with that in adult mice. Next Generation Sequencing (NGS) analysis of whole liver extracts revealed approximately 40% base-editing efficiency in adults and up to ~60% efficiency in newborns, with a broader range in efficiencies (FIG. 8A). Bystander editing remained low in adults and newborns. (FIG. 8A).

Newborn mice homozygous for huR83C were treated with lipid nanoparticles (LNP) containing guide RNA and mRNA encoding ABE. It was found that the treated mice survived and grew normally to 3 weeks of age, without hypoglycemia-induced seizures, in the absence of glucose therapy. The treated homozygous huR83C mice displayed editing efficiencies up to ~60% in total liver extracts, consistent with littermate controls that were heterozygous for huR83C (FIG. 8B). It was thus demonstrated that LNP-mediated R83C correction was associated with the survival of the homozygous huR83C mice.

Reversal of GSD-1a pathology via base-editing for correction of R83C in vivo

At 3 weeks, it was validated and confirmed that the treated homozygous huR83C mice displayed proper metabolic function, with restoration of near-normal serum metabolites, including glucose, triglycerides, cholesterol, lactate, and uric acid, as demonstrated by the darker-color bars in FIG. 9A, compared to controls. Moreover, the results of biochemical assays of G6PC activity (as assessed biochemically and via lead-phosphate staining) in LNP-treated homozygous huR83C mice were consistent with those of litter-mate controls. (FIG. 9A).

Hepatomegaly is another clinical presentation of GSD1a and is primarily caused by excess glycogen and lipid deposition in the liver. To evaluate the extent of hepatomegaly in homozygous hu*G6PC*-R83C mice post base-editing, liver sections were collected from 3wk old newborn mice and immune-histochemical analysis were conducted via hematoxylin and eosin (H&E) and Oil red O staining (FIG. 9B). Significant lipid deposition (heavy H&E staining) and enlarged hepatocytes was visualized in liver sections from homozygous mice exhibiting negligible G6Pase activity (FIG. 9B, center panels, H&E), consistent with GSD-1a. In the case of base-edited homozygous huG6PC-R83C mice showing restored G6PC activity ("HOM huR83C", right panels, FIG. 9B), lipid deposition was significantly reduced and consistent with controls (left panel), (FIG. 9B, Lipid), and restoration of hepatocyte size was apparent. Accordingly, the immuno-histochemical analyses revealed normal hepatocyte size and lipid deposition in LNP-treated mice. (FIG. 9B). Taken together, the data

demonstrate the ability of base-editing to correct the R83C mutation and to reverse the metabolic defects and pathology associated with GSD1a. In addition, these data lend further support of the functional restoration and positive clinical outcomes via base-editing for GsD-1a.

As described in this Example, novel adenine base editors and guide RNA that achieved precise correction of R83C *in vitro* and *in vivo* were generated and validated. LNP-mediated delivery of ABE and gRNA yielded significant base-editing efficiency, namely, up to ~60% base editing efficiency, with restoration of hepatic G6Pase activity and metabolic function consistent with controls.

Single LNP dose administration maintains euglycemia during a 24 hour fasting challenge via base editing

A hallmark symptom of GSD-1a pathology is fasting hypoglycemia, with a precipitous decline in blood glucose levels within minutes. A full proof-of-concept study was conducted in GSD-1a transgenic mice, homozygous for hu*G6PC*-R83C, to test whether the animals could sustain a 24 hour (hr) fast after base-editing treatment as described herein. In this study, 100% animal survival was achieved post-24hr fasting period in LNP-treated (1.5mpk) GSD-1a animals and in healthy controls. In addition, normal fasting glucose levels were measured in control mice and in treated mice pre- and post-24hr fasting, which maintained levels above hypoglycemic therapeutic threshold (>60mg/dL), (FIG. 10).

Kaplan-Meier survival estimates for homozygous huG6PC-R83C mice

Kaplan-Meier survival curves were generated to estimate the survival of newborn transgenic mice homozygous for huG6PC-R83C, either post base-editing via ABE mRNA (ABE-treated) or untreated (Untreated), (FIG. 12) Newborn mice were genotyped via PCR analysis of genomic tail DNA using the following primers, a universal forward primer (5'-ACCTACTGATGATGCACCTTTGATCAATAGAT-3'), (SEQ ID NO: 424), a mouse specific reverse primer (5'-CATCACCCCTCGGGATGGTTCTT-3'), (SEQ ID NO: 425), a human specific reverse primer 1 (5'-CAGCCCAGAATCCCAACCACAAAAT-3'), (SEQ ID NO: 426), and a human specific reverse primer 2 (5'-AGACCAGCTCGACTTGGGATGG-3'), (SEQ ID NO: 427). Survival was noted for transgenic mice homozygous for huG6PC-R83C. Untreated mice were either still-born (n=6) or died at 8 hours (n=6) and 24 hours (n=1). Administration of 15% glucose injections extended survival of the animals to 32

hours (n=5), 48 hourrs (n=2), and 56 hours (n=2). All ABE-treated mice homozygous for huG6PC-R83C survived to the termination of study at 3 weeks.

G6PC target sequences for use with base editors to correct the R83C mutation

In addition to the G6PC target sequence and guide RNA described in Example 1, alternative G6PC target sequences that can be used in conjunction with the base editors to effect base editing to correct the R83C mutation as described herein include those shown in **Table 19**. As shown, the target sequences include the types of PAMs and base editors, such as IBEs as described herein, suitable for use. In the protospacer sequences in **Table 19**, the position of the targeted "A" nucleotide (i.e., A8-A15) is shown in bold/underline. G6PC gRNA sequences hybridize to the complement of the G6PC target sequence shown in **Table 19**. The PAM sequences (*e.g.*, SpCas9) are underlined in **Table 19**.

Inlaid base editors (IBEs) noted in **Table 19** refer to structures of Cas9 and TadA having an architecture in which the deaminase domains are internal to (embedded inside) a CRISPR-Cas protein, e.g., Cas9. The IBE architecture allows for a greater breadth of potential base editing targets compared with other base editors and is not limited by the requirement of a suitably positioned Cas9 protospacer adjacent motif sequence. Such IBEs exhibited shifted editing windows and exhibited greater editing efficiency, thus allowing for the editing of targets outside the canonical editing window with reduced DNA and RNA off-target editing frequency. Accordingly, IBEs expand the breadth of potential base editing targets by extending the range of editing windows that can be created for any given CRISPR-Cas protein used to target the DNA. Through the insertion of the deaminase into a CRISPR protein at different strategic positions, the active site of the deaminase can be repositioned, making IBEs capable of editing outside the traditional editing window. IBE architectures are described hereinabove and in S. Haihua Chu et al., *The CRISPR Journal*, Vol. 4, No. 2; published online 20 April 2021 (DOI: 10.1089/crispr.2020.0144).

Table 19. Protospacer+PAM sequences (5' to 3') for correcting the R83C mutation, where the PAM sequence is underlined

CCACCAGTATGGAC**A**CTGTC CAAA (SEQ ID NO: 416) with spCas9-NRRH A15 can use IBE architecture
CACCAGTATGGAC**A**CTGTCC AAAG (SEQ ID NO: 417) with spCas9-NRRH A14 can use IBE architecture
ACCAGTATGGAC**A**CTGTCCA AAGA (SEQ ID NO: 418) with spCas9-NRRH A13 can use IBE architecture

CCAGTATGGAC**A**CTGTCCAA AGAG (SEQ ID NO: 419) with spCas9-NRRH A12 can use IBE architecture
CAGTATGGAC**A**CTGTCCAAA GAGA (SEQ ID NO: 420) with spCas9-NRRH A11 can use IBE architecture
AGTATGGAC**A**CTGTCCAAAG AGA (SEQ ID NO: 421) with spCas9-NGA A10 can use IBE architecture
GTATGGAC**A**CTGTCCAAAGA GAAT (SEQ ID NO: 422) with spCas9-NRRH A9 can use IBE architecture
TATGGAC**A**CTGTCCAAAGAG AATC (SEQ ID NO: 423) with spCas9-NRTH A8 can use IBE architecture

The gRNA sequences which hybridize to the complement of the G6PC target sequence in **Table 19** are as follows (5' to 3'): CCACCAGUAUGGACACUGUC (SEQ ID NO: 371); CACCAGUAUGGACACUGUCC (SEQ ID NO: 372); ACCAGUAUGGACACUGUCCA (SEQ ID NO: 373); CCAGUAUGGACACUGUCCAA (SEQ ID NO: 374); CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370); AGUAUGGACACUGUCCAAAG (SEQ ID NO: 375); GUAUGGACACUGUCCAAAGA (SEQ ID NO: 376); and UAUGGACACUGUCCAAAGAG (SEQ ID NO: 377).

A protospacer and PAM sequence for use in the products, compositions and methods described herein is, (5' to 3'), CAGTATGGACACTGTCCAAAGAGAAT (SEQ ID NO: 395), in which the PAM sequence,_GAGAAT, is underlined. The gRNA sequence, as presented *supra*, which hybridizes to the complement of the target sequence is CAGUAUGGACACUGUCCAAA (3 ' PAM sequence GAGAAT as shown in the sequence above) (SEQ ID NO: 370).

The gRNA sequence used in the methods described herein comprises or consists of:
CACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAA AACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 409)
or
CCACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUA AAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 410).

In some embodiments, the gRNA sequence used in the methods described herein comprises one or more modified nucleosides. Two exemplary sequences are provided below:

sgRNA_096: 23 nt protospacer

mCsmAsmCsCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUC
UACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU
(SEQ ID NO: 409)

sgRNA_097: 34 nt protospacer

mCsmCsmAsCCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAU
CUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU
(SEQ ID NO: 410).

Example 3: Optimization via new ABE variants for R83C correction

To diversify and expand a selection of saABE variants for R83C correction, the adenosine deaminase domain (referred to as a TadA domain) was engineered via directed evolution to include a combination of mutations. Such mutations were based, at least in part, on novel molecular engineering efforts. **Table 20** provides a list of novel SaABE variants, 823-829, and shows mutations present in the TadA domain that were tested alongside two other TadA variants, MSP605 and MSP680, both *in vitro* and *in vivo*. Combinations of mutations, including Y147D, F149Y, and D167N, in the TadA deaminase yielded an increase in on-target editing and reduction in bystander editing (**FIG. 3**). By way of further example, other engineered saABE variants that demonstrated effective on-target editing and reduced bystanding editing include at least one or more of the following mutations: L36H; I76Y; V82G; Q154S; and N157K, e.g., an saABE variant containing at least a combination of I76Y; V82G; Q154S.

Table 20: Novel saABE variants for use in correcting a GSD-Ia-R83C mutation

Variant		TadA Amino Acid Number							
		36	**76**	**82**	**147**	**149**	**154**	**157**	**167**
TadA-7.10		L	I	V	Y	F	Q	N	D
605	dimer			G	T		S		
680	mono		Y	G	T		S		
823	dimer	H		G	T		S	K	
824	dimer			G	D	Y	S		N
825	dimer	H		G	D	Y	S	K	N
827	mono	H	Y	G	T		S	K	
828	mono		Y	G	D	Y	S		N
829	mono	H	Y	G	D	Y	S	K	N

Example 4: Base editor sequences

Polynucleotide (mRNA, DNA) sequences and corresponding amino acid sequence of a representative base editor as described herein are presented below.

MSP828 base editor mRNA sequence

The MSP828 base-editor mRNA open reading frame (ORF) is shown by underlining in the below mRNA sequence.

AGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACCAUGAGCGAGGUCGAG
UUCUCUCACGAAUAUUGGAUGAGACACGCUCUCACCCUGGCUAAGAGAGCCAGGGACGAAAG
AGAGGUGCCAGUUGGCGCUGUCCUGGUGUUGAACAAUCGCGUCAUCGGAGAAGGAUGGAAUC
GCGCCAUUGGCCUGCACGAUCCAACCGCACAUGCCGAAAUUAUGGCUCUGCGGCAAGGCGGC
CUCGUGAUGCAAAAUUACAGACUGUACGAUGCUACCCUCUACGGCACCUUCGAGCCCUGUGU
CAUGUGUGCUGGGGCAAUGAUUCACUCCCGGAUUGGCCGCGUGGUGUUUGGAGUGCGGAAUG
CCAAGACUGGCGCCGCUGGAUCUCUGAUGGACGUCCUGCACUAUCCUGGGAUGAACCACCGG
GUCGAGAUCACAGAGGGAAUUCUGGCUGACGAGUGCGCAGCCCUGCUGUGCGACUUCUAUAG
AAUGCCCAGAUCGGUGUUCAACGCCCAGAAAAAAGCUCAGAGCAGUACCAAUUCCGGCGGAA
GCAGCGGAGGAUCUUCUGGAAGCGAAACCCCAGGCACCAGCGAGUCUGCCACACCAGAAUCA
UCUGGCGGUAGCUCCGGCGGCUCCAAGAGAAAUUACAUCCUGGGCCUCGCCAUCGGCAUCAC
CUCUGUCGGCUACGGCAUCAUCGACUACGAGACACGGGAUGUGAUUGAUGCCGGCGUGCGGC
UGUUCAAAGAGGCCAACGUCGAGAACAACGAGGGCCGCAGAUCUAAGAGGGGAGCCAGACGG
CUGAAGAGAAGGCGGAGACACAGAAUCCAGAGAGUGAAGAAGCUGCUGUUCGACUACAACCU
GCUGACCGACCACAGCGAGCUGAGCGGCAUCAACCCUUAUGAGGCCAGAGUGAAGGGCCUGA
GCCAGAAGCUGAGCGAGGAAGAGUUUAGCGCCGCACUGCUGCACCUGGCCAAGAGAAGAGGC
GUGCACAACGUGAACGAAGUGGAAGAGGACACCGGCAACGAGCUGUCCACCAAAGAGCAGAU
CAGCAGAAACAGCAAGGCCCUGGAAGAGAAAUACGUGGCCGAACUGCAGCUGGAACGGCUGA
AAAAGGAUGGCGAAGUGCGGGGCAGCAUCAACCGGUUCAAGACCAGCGACUACGUGAAAGAA
GCCAAACAGCUGCUGAAGGUGCAGAAGGCCUACCACCAGCUGGACCAGAGCUUCAUCGACAC
CUACAUCGACCUGCUGGAAACCCGGCGGACCUACUAUGAAGGACCUGGCGAGGGAAGCCCCU
UCGGCUGGAAGGACAUCAAAGAAUGGUACGAGAUGCUGAUGGGCCACUGCACCUACUUUCCC
GAGGAACUGCGGAGCGUGAAGUACGCCUACAACGCCGACCUGUACAACGCCCUGAACGACCU
GAACAACCUCGUGAUCACCCGGGACGAGAACGAGAAGCUGGAAUAUUACGAGAAGUUCCAGA
UCAUCGAGAACGUGUUCAAGCAGAAGAAGAAGCCCACACUGAAGCAGAUCGCCAAAGAGAUC
CUGGUCAACGAGGAAGAUAUCAAGGGCUACAGAGUGACCAGCACCGGCAAGCCCGAGUUCAC
CAACCUGAAGGUGUACCACGACAUCAAGGAUAUCACAGCCCGGAAAGAGAUUAUUGAGAACG

CCGAGCUGCUGGACCAAAUCGCCAAGAUCCUGACCAUCUACCAGUCCUCCGAGGACAUCCAA

GAGGAACUGACCAAUCUGAACUCCGAGCUGACCCAAGAAGAGAUCGAGCAGAUCUCUAAUCU

GAAGGGGUACACAGGCACCCACAACCUGAGCCUGAAGGCCAUCAACCUGAUCCUGGACGAGC

UGUGGCACACCAACGACAACCAGAUUGCCAUCUUCAACCGGCUGAAGCUGGUGCCCAAGAAG

GUGGACCUCAGCCAGCAGAAAGAAAUCCCCACCACACUGGUGGACGACUUCAUUCUGAGCCC

CGUGGUCAAGAGAAGCUUCAUCCAGAGCAUCAAAGUGAUCAACGCCAUCAUCAAGAAGUACG

GGCUGCCCAACGAUAUCAUCAUCGAGCUGGCCCGCGAGAAGAACUCCAAGGACGCUCAGAAA

AUGAUCAACGAGAUGCAGAAGCGGAACCGGCAGACCAACGAGCGGAUCGAGGAAAUCAUCCG

GACCACCGGCAAAGAGAACGCCAAGUACCUGAUCGAGAAGAUCAAGCUGCACGACAUGCAAG

AGGGCAAGUGUCUGUACAGCCUGGAAGCCAUUCCUCUGGAAGAUCUGCUGAACAAUCCCUUC

AACUACGAGGUGGACCACAUCAUCCCCAGAAGCGUGUCCUUCGACAACAGCUUCAACAACAA

GGUGCUCGUGAAGCAAGAGGAAAACUCCAAGAAGGGCAACAGAACCCCCAUUCCAGUACCUGA

GCAGCUCCGACAGCAAGAUCAGCUACGAAACCUUCAAGAAGCACAUCCUGAAUCUGGCCAAA

GGCAAGGGCCGCAUCAGCAAGACCAAGAAAGAAUACCUGCUCGAGGAACGGGACAUCAACAG

AUUCAGCGUGCAGAAAGACUUCAUCAAUCGGAACCUGGUGGACACCAGAUACGCCACCAGAG

GCCUGAUGAAUCUGCUGAGAAGCUACUUCCGCGUGAACAAUCUGGACGUGAAAGUCAAGUCC

AUCAACGGCGGCUUCACCAGCUUUCUGCGGAGAAAGUGGAAGUUCAAGAAAGAGCGGAACAA

GGGCUAUAAGCACCACGCCGAGGACGCCCUGAUCAUUGCCAACGCCGAUUUCAUCUUCAAAG

AGUGGAAGAAACUGGACAAGGCCAAAAAAGUGAUGGAAAACCAGAUGUUCGAGGAAAAGCAG

GCCGAGAGCAUGCCCGAGAUCGAAACCGAGCAAGAGUACAAAGAGAUUUUCAUCACGCCCCA

CCAGAUCAAGCACAUUAAGGACUUCAAGGACUACAAGUACAGCCACCGCGUGGACAAGAAGC

CUAAUAGAGAGCUGAUUAACGACACCCUGUACAGCACCCGGAAGGACGACAAGGGCAAUACC

CUGAUCGUCAACAACCUGAACGGCCUGUACGACAAGGACAACGACAAGCUCAAGAAGCUGAU

CAACAAGAGCCCCGAGAAACUGCUGAUGUACCACCACGAUCCUCAGACCUACCAGAAACUGA

AGCUCAUCAUGGAACAGUACGGCGACGAGAAGAAUCCCCUGUACAAGUACUACGAGGAAACC

GGGAACUACCUGACCAAGUACUCCAAAAAAGGACAAUGGGCCCGUGAUCAAGAAGAUUAAGUA

UUACGGCAACAAGCUGAAUGCCCACCUGGACAUCACCGACGACUACCCCAACUCCAGAAACA

AGGUGGUCAAGCUGUCCCUGAAGCCUUACAGAUUCGACGUGUACCUGGACAACGGCGUGUAC

AAGUUCGUGACCGUGAAGAACCUGGAUGUGAUCAAAAAAGAAAACUACUACGAAGUGAACAG

CAAGUGCUAUGAGGAAGCCAAGAAACUCAAGAAAAUCAGCAACCAGGCCGAGUUUAUCGCCU

CCUUCUACAACAACGAUCUGAUCAAGAUCAACGGGGAGCUGUAUAGAGUGAUUGGGGUCAAC

AAUGACCUGCUGAACCGGAUCGAAGUCAACAUGAUCGACAUCACCUACCGCGAGUACCUCGA

GAACAUGAACGACAAGAGGCCUCCACGGAUCAUUAAGACAAUCGCCAGCAAGACGCAGAGCA

UUAAGAAGUACAGCACUGACAUUCUGGGCAACCUGUACGAAGUCAAGAGCAAAAAGCACCCG

CAGAUUAUCAAGAAAGGCGAGGGCGCCGACAAGAGAACAGCCGAUGGUUCCGAGUUCGAAAG
CCCCAAGAAGAAGAGGAAAGUCUAGUUAAUUAAGCUGCCUUCUGCGGGGCUUGCCUUCUGGC
CAUGCCCUUCUUCUCUCCCUUGCACCUGUACCUCUUGGUCUUUGAAUAAAGCCUGAGUAGGA
AGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

(SEQ ID NO: 396)

MSP828 base editor DNA sequence

ATGAGCGAGGTCGAGTTCTCTCACGAATATTGGATGAGACACGCTCTCACCCTGGCTAAGAG
AGCCAGGGACGAAAGAGAGGTGCCAGTTGGCGCTGTCCTGGTGTTGAACAATCGCGTCATCG
GAGAAGGATGGAATCGCGCCATTGGCCTGCACGATCCAACCGCACATGCCGAAATTATGGCT
CTGCGGCAAGGCGGCCTCGTGATGCAAAATTACAGACTGTACGATGCTACCCTCTACGGCAC
CTTCGAGCCCTGTGTCATGTGTGCTGGGGCAATGATTCACTCCCGGATTGGCCGCGTGGTGT
TTGGAGTGCGGAATGCCAAGACTGGCGCCGCTGGATCTCTGATGGACGTCCTGCACTATCCT
GGGATGAACCACCGGGTCGAGATCACAGAGGGAATTCTGGCTGACGAGTGCGCaGCCCTGCT
GTGCgacTTCTaTAGAATGCCCAGAtcGGTGTTCAACGCCCAGAAAAAAGCTCAGAGCAGtA
CCaATTCCGGCGGAAGCAGCGGAGGATCTTCTGGAAGCGAAACCCCAGGCACCAGCGAGTCT
GCCACACCAGAATCATCTGGCGGTAGCTCCGGCGGCTCCAAGAGAAATTACATCCTGGGCCT
CGCCATCGGCATCACCTCTGTCGGCTACGGCATCATCGACTACGAGACACGGGATGTGATTG
ATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTCGAGAACAACGAGGGCCGCAGATCTAAG
AGGGGAGCCAGACGGCTGAAGAGAAGGCGGAGACACAGAATCCAGAGAGTGAAGAAGCTGCT
GTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCTTATGAGGCCA
GAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTTAGCGCCGCACTGCTGCACCTG
GCCAAGAGAAGAGGCGTGCACAACGTGAACGAAGTGGAAGAGGACACCGGCAACGAGCTGTC
CACCAAAGAGCAGATCAGCAGAAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGC
AGCTGGAACGGCTGAAAAAGGATGGCGAAGTGCGGGGCAGCATCAACCGGTTCAAGACCAGC
GACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCA
GAGCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAAGGACCTG
GCGAGGGAAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCAC
TGCACCTACTTTCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAA
CGCCCTGAACGACCTGAACAACCTCGTGATCACCCGGGACGAGAACGAGAAGCTGGAATATT
ACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACACTGAAGCAG
ATCGCCAAAGAGATCCTGGTCAACGAGGAAGATATCAAGGGCTACAGAGTGACCAGCACCGG
CAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGATATCACAGCCCGGAAAG

AGATTATTGAGAACGCCGAGCTGCTGGACCAAATCGCCAAGATCCTGACCATCTACCAGTCC
TCCGAGGACATCCAAGAGGAACTGACCAATCTGAACTCCGAGCTGACCCAAGAAGAGATCGA
GCAGATCTCTAATCTGAAGGGGTACACAGGCACCCACAACCTGAGCCTGAAGGCCATCAACC
TGATCCTGGACGAGCTGTGGCACACCAACGACAACCAGATTGCCATCTTCAACCGGCTGAAG
CTGGTGCCCAAGAAGGTGGACCTCAGCCAGCAGAAAGAAATCCCCACCACACTGGTGGACGA
CTTCATTCTGAGCCCCGTGGTCAAGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCA
TCATCAAGAAGTACGGGCTGCCCAACGATATCATCATCGAGCTGGCCCGCGAGAAGAACTCC
AAGGACGCTCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGGAT
CGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGC
TGCACGACATGCAAGAGGGCAAGTGTCTGTACAGCCTGGAAGCCATTCCTCTGGAAGATCTG
CTGAACAATCCCTTCAACTACGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAA
CAGCTTCAACAACAAGGTGCTCGTGAAGCAAGAGGAAAACTCCAAGAAGGGCAACAGAACCC
CATTCCAGTACCTGAGCAGCTCCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATC
CTGAATCTGGCCAAAGGCAAGGGCCGCATCAGCAAGACCAAGAAAGAATACCTGCTCGAGGA
ACGGGACATCAACAGATTCAGCGTGCAGAAAGACTTCATCAATCGGAACCTGGTGGACACCA
GATACGCCACCAGAGGCCTGATGAATCTGCTGAGAAGCTACTTCCGCGTGAACAATCTGGAC
GTGAAAGTCAAGTCCATCAACGGCGGCTTCACCAGCTTTCTGCGGAGAAAGTGGAAGTTCAA
GAAAGAGCGGAACAAGGGCTATAAGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCG
ATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATG
TTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAAGAGTACAAAGAGAT
TTTCATCACGCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACC
GCGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACAGCACCCGGAAGGAC
GACAAGGGCAATACCCTGATCGTCAACAACCTGAACGGCCTGTACGACAAGGACAACGACAA
GCTCAAGAAGCTGATCAACAAGAGCCCCGAGAAACTGCTGATGTACCACCACGATCCTCAGA
CCTACCAGAAACTGAAGCTCATCATGGAACAGTACGGCGACGAGAAGAATCCCCTGTACAAG
TACTACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACAATGGGCCCGTGAT
CAAGAAGATTAAGTATTACGGCAACAAGCTGAATGCCCACCTGGACATCACCGACGACTACC
CCAACTCCAGAAACAAGGTGGTCAAGCTGTCCCTGAAGCCTTACAGATTCGACGTGTACCTG
GACAACGGCGTGTACAAGTTCGTGACCGTGAAGAACCTGGATGTGATCAAAAAAGAAAACTA
CTACGAAGTGAACAGCAAGTGCTATGAGGAAGCCAAGAAACTCAAGAAAATCAGCAACCAGG
CCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGGGAGCTGTATAGA
GTGATTGGGGTCAACAATGACCTGCTGAACCGGATCGAAGTCAACATGATCGACATCACCTA
CCGCGAGTACCTCGAGAACATGAACGACAAGAGGCCTCCACGGATCATTAAGACAATCGCCA
GCAAGACGCAGAGCATTAAGAAGTACAGCACTGACATTCTGGGCAACCTGTACGAAGTCAAG

AGCAAAAAGCACCCGCAGATTATCAAGAAAGGCgagGGCGCCGACAAGAGAACAgccgatgg
ttccgagttcgaaagccccaagaagaagaggaaagtctaG (SEQ ID NO: 397)

Translated protein (amino acid) sequence of the MSP828 base editor

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYGTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCDFYRMPRSVFNAQKKAQSSTN**SGGSSGGSSGSETPGTSES**
ATPESSGGSSGGSKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK
RGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHL
AKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTS
DYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGH
CTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQ
IAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQS
SEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLK
LVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNS
KDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDL
LNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHI
LNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLD
VKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQM
FEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKD
DKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYK
YYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYL
DNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYR
VIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVK
SKKHPQIIKKGEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 398)

In the above MSP828 protein (amino acid) sequence, the TadA 7.10 domain is indicated by underlining; the linker sequence is indicated by bold font; the SaCas9 sequence is indicated by double underlining; and the nuclear localization signal (bpNLS) is indicated by dotted line underlining.

Example 5: Heavily modified SaCas9 gRNA sequences

This example provides representative saCas9 guide RNA sequences designed to include numerous sequence modifications (heavily modified SaCas9 gRNA sequences or "heavy mods").

Chemical modifications to RNAs, e.g., gRNAs, sgRNAs, provide protection to the RNA molecule and allow for efficient genomic and base editing. Chemically protected gRNAs can enhance stability and editing efficiency *in vivo* and *ex vivo*. Non-limiting examples of suitable chemical modifications include 2′-*O*-methyl (2′-OMe), phosphorothioate (PS), 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), and constrained ethyl (S-cEt), which may be employed in the synthesis of sgRNAs. In a particular embodiment, the chemical modifications are 2′-*O*-methyl (2′-OMe) modifications. The modified guide RNAs may improve saCas9 efficacy and also specificity. The effect of an individual modification varies based on the position and combination of chemical modifications used as well as the inter- and intramolecular interactions with other modified nucleotides. By way of example, S-cEt has been used to improve oligonucleotide intramolecular folding.

Heavy mod designs currently do not exist for Cas orthologs such as saCas9, Cas12, etc., e.g., those which are not SpCas9. Such SpCas9 heavy mods can increase base editing ~2 fold *in vivo*. For such modifications, mN = 2′-OMe; Ns = phosphorothioate (PS), where "N" represents the term nucleotide, as would be understood by one having skill in the art. In some cases, a nucleotide (N) may contain two modifications, for example, both a 2′-OMe and a PS modification. Representative heavy mods gRNAs (heavy mod series 1 gRNAs for R83C (saCas9), as shown in (1) below) were utilized in experiments in which adult transgenic mice heterozygous for huG6PC-R83C were dosed at a sub-saturating dose of 1mpk of 1:1 ratio of gRNA:editor mRNA (FIG. 11). The heavy mods as described here provide a new universal approach to increase *in vivo* base editing that is independent of sequence.

Provided below are examples of heavy mod gRNA sequences and end mod gRNA sequences. In particular, shown below are sequences in connection with the following: (1) heavy mods series 1; (2) heavy mods series 2; (3) protospacer extension; (4) direct repeat/anti-repeat lengths/sequence context, and the "base" sequences (end mods); and (5) gRNA sequences (end mod sgRNA sequences) designed for correcting the huG6PC-R83C mutation as described and exemplified *supra*. As will be appreciated by those skilled in the

art, Cas9 evolved to use a trans-activating RNA (tracrRNA) that base pairs via an anti-repeat sequence to the repeat sequence of individual crRNAs. A more commonly used single gRNA constitutes a fusion of the tracrRNA and crRNA through a short RNA loop between the repeat-anti-repeat sequences in the Upper Stem (G. Mullally et al., 2020, *Nucleic Acids Research*, 48(12):6811-6823; M. Jinek et al., 2012, *Science*, 337:816-821). The direct repeat regions contain sequences for processing pre-crRNA into mature crRNA and tracrRNA binding. The "direct repeat region" combined with the tracrRNA forms the scaffold portion of a gRNA and the "spacer region" forms the target sequence.

In each of the following sequences, mN = 2'OMe; Ns = phosphorothioate, where "N" designates any nucleotide or nucleobase. In the below sequences, "s" indicates that the preceding nucleotide possesses a 3' phosphorothioate, and the "m" indicates the following nucleotide is a 2' OMe. A nucleotide with a phosphorothioate and 2' OMe is denoted as "mNs;" when there are two modifications next to each other, the notation is "mNsmNs."

1. Heavy mods series 1

gRNA578: 50% modified with 2'OMe

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUmAmGmUACUCmUGmUmAmAmUGmAA AmAmUmUmACmAGAAUCUACmUmAAAACAAGGCAAmAAUGmCCmGUGUmUmUmAmUmCmUmC mGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404)

gRNA579: 50% modified, 5' extended with 3 Us (uridines)

mUsmUsmUsCAGmUAUmGmGmACAmCUGUCCAAAmGUUUUmAmGmUACUCmUGmUmAmAmUG mAAAmAmUmUmACmAGAAUCUACmUmAAAACAAGGCAAmAAUGmCCmGUGUmUmUmAmUmCm UmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 405)

gRNA580: 66% modified

mCsmAsGsmUAUmGmGmAmCAmCmUGUCCAAmAmGUmUUmUmAmGmUACUmCmUmGmUmAmA mUGmAmAmAmAmUmUmACmAmGAAmUCUACmUmAmAAACAAGmGCAAmAAUGmCmCmGUGmU mUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404)

gRNA581: 69% modified

mCsmAsGsmUmAUmGmGmAmCAmCUGUCCAAmAmGUUUmUAmGmUACUmCmUmGmUmAmAmU mGmAmAmAmAmUmUmAmCmAmGmAAmUCUACUmAmAAACAAmGmGmCmAmAmAAUmGmCmCG UGmUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmU smU (SEQ ID NO: 404)

gRNA582: 84% modified

mCsmAsGsmUmAUmGmGmAmCAmCmUGUCmCmAAmAmGmUmUmUmUmAmGmUAmCmUmCmUm
GmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAmAmUCUACmUmAmAmAAmCAmAmGmGmCm
AmAmAAUmGmCmCmGmUGmUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmC
mGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404)

2. Heavy mods series 2

gRNA964: 108 nt (longer DR/anti-DR); 24% modified; hairpin (DR/anti-DR) mods only

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGmUmAmAmUmGmAmAmAmA
mUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUG
GCGAGAUsmUsmUsmU (SEQ ID NO: 404)

gRNA965: 108 nt (longer DR/anti-DR); 48% modified; hairpin (DR/anti-DR) mods + 3' mods

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGmUmAmAmUmGmAmAmAmA
mUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCGUGUmUmUmAmUmCmUmCmGmUm
CmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404)

gRNA966: 108 nt (longer DR/anti-DR); 53% modified; hairpin (DR/anti-DR) mods + 3' mods + protospacer mods

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUAGUACmUmCmUmGmUmAmAmUmGmA
mAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCmGUGUmUmUAmUmCmUm
CmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAUsmUsmUsmU (SEQ ID NO: 404)

gRNA967: 100 nt (shorter DR/anti-DR); 43% modified; hairpin (DR/anti-DR) mods + 3' mods

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGGmAmAmAmCmAmGmAAUC
UACUAAAACAAGGCAAAAUGCCGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmU
mGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 406)

gRNA968: 100 nt (shorter DR/anti-DR); 49% modified; hairpin (DR/anti-DR) mods + 3' mods + protospacer mods

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUAGUACmUmCmUmGmGmAmAmAmCmA
mGmAAUCUACUAAAACAAGGCAAAAUGCCmGUGUmUmUAmUmCmUmCmGmUmCmAmAmCmUm
UmGmUmUmGmGmCmGmAmGmAUsmUsmUsmU (SEQ ID NO: 406)

3. Variable length protospacers and sgRNAs

sgRNA_094: 21 nt protospacer

mCsmCsmAsGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 407)

sgRNA_095: 22 nt protospacer

mAsmCsmCsAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCU ACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 408)

sgRNA_096: 23 nt protospacer

mCsmAsmCsCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUC UACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 409)

sgRNA_097: 34 nt protospacer

mCsmCsmAsCCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAU CUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 410)

4. Modified direct repeat / anti-direct repeat sequences and gRNAs

gRNA868: R83C sgRNA_029 16bp DR; 104 nt

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAGAAAUACAGAAUCUACUAAA ACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 411)

gRNA869: R83C sgRNA_029 16bp CG DR; 104 nt

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGCGGAAACGCAGAAUCUACUAAA ACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 412)

gRNA870: R83C sgRNA_029 14bp DR ran et al; 100 nt

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAA GGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 406)

gRNA871: R83C sgRNA_029 12bp DR; 96 nt

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCGAAAGAAUCUACUAAAACAAGGCA AAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 413)

gRNA872: R83C sgRNA_029 12bp CC DR; 96 nt

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACCCGAAAGCAUCUACUAAAACAAGGCA AAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 414)

5. End mod sgRNA sequences designed for use in correcting the huG6PC R83C mutation

sgRNA_029

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUAC UAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 415)

sgRNA_093

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUAC UAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 415)

Example 6: Materials and Methods

Materials and methods utilized in the examples and experiments therein as described *supra* are set forth below.

Animal care

All animal studies were conducted under Taconic's Excluded Flora health standard. To sustain survival of hu*G6PC*-R83C mice, a glucose therapy consisting of daily administered subcutaneous injections of 100-150ul of 15% glucose per mouse. Glucose injections were not administered to mice post LNP treatment with base-editor mRNA and gRNA.

***In vivo* LNP-dosing work-flow**

To correct the p.R83C mutation in the hu*G6PC*-R83C homozygous mice, LNP co-formulations of base-editor mRNA and gRNA were administered at a 1.5 mpk (milligram per kilogram) dose via the temporal vein of mice at age P1, shortly post birth. Glucose therapy was not administered to LNP-treated mice. LNP-treated mice continued to be cared for alongside littermate controls by the respective birth mother until weaning (21 days), at which point they were phenotyped. For all studies, age matched wild-type and heterozygous

hu*G6PC*-R83C littermates were used as controls. At day 21, genomic DNA harvested from livers, growth characteristics, and serum and liver markers were analyzed.

Lipid Nanoparticle (LNP) Formulations

The base editor (mRNA encoding the base editor) and guide RNA were co-encapsulated at a 1:1 weight ratio in a lipid nanoparticle. The LNPs were generated by rapidly mixing an aqueous solution of the RNA at a pH of 3.0 with an ethanol solution containing four lipid components: an ionizable lipid, DSPC, cholesterol, and a lipid-anchored PEG. The two solutions were mixed using the benchtop microfluidics device from Precision Nanosystems. Post mixing, the formulations were dialyzed overnight at 4°C against 1x TBS (Sigma-Aldrich, catalog #94158). They were subsequently concentrated down using 100K MWCO Amicon Ultra centrifugation tubes (Millipore Sigma, catalog# UFC910096), and filtered with 0.2 micron filters (Pall corperation, Catalog #4602). Total RNA concentration of the was determined using Quant-iT Ribogreen (ThermoFisher Scientific, catalog# R11491); particle size was determined by using the Malvern Panalytical Zetasizer.

Next Generation Sequencing (NGS)

Next generation sequencing (NGS) was used to determine the frequency of base-edited alleles in genomic DNA from whole liver extracts of LNP-treated animals. Following LNP-treatment, mice were euthanized, and the entire liver was removed and snap frozen in liquid nitrogen. Frozen mouse livers were ground to a powder form using Geno/Grinder 2010 (Ops Diagnostics, Lebanon, NJ, USA), and genomic DNA was isolated from the liver powder using Quick Extract lysis buffer according to manufacture's specifications. Genomic DNA was directly used in subsequent PCR amplification steps to produce a ~170-nucleotide fragment harboring hu*G6PC* exon 2 using the primer pair: Forward primer, GGGCATTAAACTCCTTTGGG (SEQ ID NO: 393) and reverse primer, AGTCTCACAGGTTACAGGGA (SEQ ID NO: 394). NGS adapters were added, and the resulting amplicons were sequenced using an Illumina MiSeq instrument according to the manufacturer's instructions.

Serum Metabolites

To measure serum metabolites, blood was collected from R83C humanized transgenic mice. Serum was then separated and extracted from whole blood, which was subsequently

used for metabolite assays. For a relevant and comprehensive post-study assessment, serum glucose, serum cholesterol, and serum triglycerides were all analyzed. Serum glucose and serum cholesterol were measured using ThermoFisher Scientific (Waltham, MA, USA) Infinity Glucose Liquid Stable Reagent (Cat#: TR15421) and Infinity Cholesterol Liquid Stable Reagent (Cat#: TR13421), respectively. Serum triglycerides were measured using the Serum Triglyceride Quantification Kit (Cat#: MAK266) from Sigma-Aldrich (St. Louis, MO, USA). Uric acid was measured using the Uric Acid Liquid Stable Reagent per manufacturer specifications (Thermo Fisher Scientific (Waltham, MA, USA). Serum lactate was analyzed using the EnzyFluo L-Lactate Assay Kit from BioAssay Systems (Hayward, CA, USA).

Fasting blood glucose analysis of mice involved blood sampling via the tail vein pre- and post-24hours after food deprivation. Blood glucose levels were measured using the HemoCue Glucose 201 System (HemoCue America, CA, USA).

Glucose-6-Phosphatase-alpha activity assay

Liver microsome isolation and microsomal phosphohydrolase assays were performed as described by Lei, K.-J., et al., 1996, Nature Genetics, 13(2):203-9. Assay methodology in Arnaotova et al. (2021, Mol. Therapy., Vol. 29, No 4) is described as follows: "Glucose-6-phosphatase dependent substrate transport in the glycogen storage disease type-1a mouse. Nat. Genet. 13, 203-209). In phosphoydrolase assays, reaction mixtures (50uL) containing 50mM sodium cacodylate buffer (pH 6.5), 2mM EDTA, 10mM Glucose-6-phosphate (G6P), and appropriate amounts of microsomal preparations were incubated at 30°C for 10 minutes. Disrupted microsomal membranes were prepared by incubating intact membranes in 0.2% deoxycholate for 20 minutes at 4°C. Non-specific phosphatase activity was estimated by pre-incubating disrupted microsomal preparations at pH 5 for 10 minutes at 37°C to inactivate the acid-labile G6Pase-alpha. One unit of G6Pase-alpha activity represents one nmol G6P hydrolysis per minute per mg microsomal protein. The lower level of quantitation for the microsomal G6Pase-alpha assay is 2 units."

Enzyme histochemical analysis of G6Pase-alpha was performed as described in Lee, Y.M., Jun, H.S. Pan, C.-J. Lin, S.R., Wilson, L.H., Mansfield, B.C., and Chou, J.Y. (2012). Prevention of hepatocyellular adenoma and correction of metabolic abnormalities in murine glycogen storage disease type Ia by gene therapy. Hepatology 56, 1719-1729. As described in Arnaotova et al., (2021, Mol. Therapy., Vol. 29, No 4), 10um-thick liver tissue sections were incubated for 10min at room temperature in a solution containin 40mM Tris-maleate

(pH 6.5), 10mM G6P, 300 mM sucrose, and 3.6mM lead nitrate. After rinsing, liver sections were incubated for 2min at room temperature in 0.09% ammonium sulfide solution, and the trapped lead phosphate was visualized following conversion to the brown-colored lead sulfide.

Immunohistochemistry

Immunohistochemical procedures were performed as described in Arnaotova et al., 2021, Mol. Therapy., Vol. 29, No 4. In brief, H&E staining was performed on liver sections preserved in 10% neutral buffered formalin, and Oil Red O staining was performed on cryopreserved optimal cutting temperature compound (OCT) embedded liver sections following standard procedures. The stained sections were visualized using the Imager A2m microscope with Axiocam 506 camera and Zen 2.6 software (Carl Zeiss, White Plains, NY, USA).

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the embodiments as described herein to be adopted to various usages and conditions. Such embodiments are also within the scope of the following claims.

The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. Absent any indication otherwise, publications, patents, and patent applications mentioned in this specification are incorporated herein by reference in their entireties.

CLAIMS

What is claimed is:

1. An adenosine deaminase variant comprising a glycine (G) at amino acid position 82, a threonine (T) or an aspartic acid (D) at amino acid position 147, a serine (S) at amino acid position 154, and one or more of a histidine (H) at amino acid position 36, a tyrosine at amino acid position 76, a tyrosine at amino acid position 149, a lysine (K) at amino acid position 157, and an asparagine (N) at amino acid position 167 of the following amino acid sequence, wherein the adenosine deaminase has at least about 85% identity to said amino acid sequence:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase.

2. An adenosine deaminase variant comprising any of the following combinations of alterations

a) I76Y + V82G + Y147T + Q154S;

b) L36H + V82G + Y147T + Q154S + N157K;

c) V82G + Y147D + F149Y + Q154S + D167N;

d) L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N;

e) L36H + I76Y + V82G + Y147T + Q154S + N157K;

f) I76Y + V82G + Y147D + F149Y + Q154S + D167N;

g) Y147D + F149Y + D167N;

h) L36H; I76Y; V82G; Q154S; and N157K;

i) I76Y; V82G; Q154S; or

j) L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N with reference to SEQ ID NO: 1:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding combinations of alterations in another adenosine deaminase.

3. The adenosine deaminase variant of claim 1 or 2, comprising the following combination of alterations I76Y + V82G + Y147D + F149Y + Q154S + D167N of SEQ ID NO: 1, or corresponding alterations in another adenosine deaminase.

4. The adenosine deaminase variant of claim 1 or 2, wherein the adenosine deaminase has at least about 90% identity to SEQ ID NO: 1.

5. The adenosine deaminase variant of claim 1 or 2, wherein the adenosine deaminase has at least about 95% identity to SEQ ID NO: 1.

6. The adenosine deaminase variant of claim 1 or 2, wherein the adenosine deaminase comprises or consists essentially of SEQ ID NO: 1.

7. A fusion protein or complex comprising a polynucleotide programmable DNA binding domain and at least one adenosine deaminase variant domain, wherein the adenosine deaminase variant domain comprises a glycine (G) at amino acid position 82, a threonine (T) or an aspartic acid (D) at amino acid position 147, a serine (S) at amino acid position 154, and one or more of a histidine (H) at amino acid position 36, a tyrosine at amino acid position 76, a tyrosine at amino acid position 149, a lysine (K) at amino acid position 157, and an asparagine (N) at amino acid position 167 of the following amino acid sequence, wherein the adenosine deaminase has at least about 85% identity to said amino acid sequence MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase.

8. The fusion protein or complex of claim 7, wherein the adenosine deaminase variant domain has at least about 90% identity to SEQ ID NO: 1.

9. The fusion protein or complex of claim 7, wherein the adenosine deaminase variant domain at least about 95% identity to SEQ ID NO: 1.

10. The fusion protein or complex of claim 7, wherein the adenosine deaminase variant domain comprises or consists essentially of SEQ ID NO: 1.

11. A fusion protein or complex comprising a polynucleotide programmable DNA binding domain and at least one adenosine deaminase variant domain, wherein the adenosine deaminase variant domain comprises any of the following combinations of alterations

a) I76Y + V82G + Y147T + Q154S;

b) L36H + V82G + Y147T + Q154S + N157K;

c) V82G + Y147D + F149Y + Q154S + D167N;

d) L36H + V82G + Y147D + F149Y + Q154S + N157K + D167N;

e) L36H + I76Y + V82G + Y147T + Q154S + N157K;

f) I76Y + V82G + Y147D + F149Y + Q154S + D167N;

g) Y147D + F149Y + D167N;

h) L36H; I76Y; V82G; Q154S; and N157K;

i) I76Y; V82G; Q154S; or

j) L36H + I76Y + V82G + Y147D + F149Y + Q154S + N157K + D167N

with reference to SEQ ID NO: 1:

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding combinations of alterations in another adenosine deaminase.

12. The fusion protein or complex of claim 7 or 11, wherein the adenosine deaminase variant domain comprises the following combination of alterations I76Y + V82G + Y147D + F149Y + Q154S + D167N of SEQ ID NO: 1, or corresponding alterations in another adenosine deaminase.

13. The fusion protein or complex of any one of claims 7-12, wherein the fusion protein comprises one adenosine deaminase variant domain.

14. The fusion protein or complex of any one of claims 7-12, wherein the fusion protein comprises a wild-type adenosine deaminase domain and an adenosine deaminase variant domain.

15. The fusion protein or complex of any one of claims 7-12, wherein the fusion protein comprises a TadA*7.10 adenosine deaminase domain and an adenosine deaminase variant domain.

16. The fusion protein or complex of any one of claims 7-15, wherein the polynucleotide programmable DNA binding domain is a Cas9 domain.

17. The fusion protein or complex of claim 16, wherein the Cas9 domain comprises a nuclease dead Cas9 (dCas9), a Cas9 nickase (nCas9), or a nuclease active Cas9.

18. The fusion protein or complex of any one of claims 7-17, wherein the polynucleotide programmable DNA binding domain is a *Staphylococcus aureus* Cas9 (SaCas9), *Streptococcus thermophilus 1* Cas9 (St1Cas9), a *Streptococcus pyogenes* Cas9 (SpCas9), or variants thereof.

19. The fusion protein or complex of any one of claims 7-17, wherein the polynucleotide programmable DNA binding domain comprises a modified SaCas9 having an altered protospacer-adjacent motif (PAM) specificity.

20. The fusion protein or complex of claim 19, wherein SaCas9 has protospacer-adjacent motif (PAM) specificity for the nucleic acid sequence 5'-NNGRRT-3'.

21. The fusion protein or complex of claim 20, wherein the SaCas9 has specificity for the nucleic acid sequence 5'-GAGAAT-3'.

22. The fusion protein or complex of any one of claims 18-21, wherein the SaCas9 is a nuclease active SaCas9, a nuclease inactive SaCas9 (SaCas9d), or a SaCas9 nickase (SaCas9n).

23. The fusion protein or complex of claim 22, wherein the SaCas9 is a nickase comprising an amino acid substitution N579A or a corresponding amino acid substitution thereof.

24. The fusion protein or complex of claim 18, comprising *Streptococcus pyogenes* Cas9 (SpCas9) or a variant thereof.

25. The fusion protein or complex of any one of claims 7-24, wherein the adenosine deaminase variant is capable of deaminating adenine in deoxyribonucleic acid (DNA).

26. The fusion protein or complex of any one of claims 7-24, comprising a linker between the polynucleotide programmable DNA binding domain and the adenosine deaminase variant domain.

27. The fusion protein or complex of claim 26, wherein the linker comprises the amino acid sequence: SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 359).

28. The fusion protein or complex of any one of claims 7-27, comprising one or more nuclear localization signals.

29. The fusion protein or complex of claim 28, wherein the nuclear localization signal is a bipartite nuclear localization signal.

30. The complex of any one of claims 7-28, wherein the polynucleotide programmable DNA binding domain non-covalently associates with the deaminase.

31. A base editor system comprising the fusion protein or complex of any one of claims 7-30, and one or more guide polynucleotides.

32. The base editor system of claim 31, wherein the one or more guide polynucleotides target the fusion protein to effect an A•T to G•C alteration of a single nucleotide polymorphism (SNP) associated with a genetic disease.

33. The base editor system of claim 32, wherein the genetic disease is Glycogen Storage Disease Type 1a (GSD1a).

34. The base editor system of any one of claims 30-32, wherein the guide polynucleotide comprises ribonucleic acid (RNA), or deoxyribonucleic acid (DNA).

35. The base editor system of any one of claims 31-34, wherein the guide polynucleotide comprises a nucleic acid sequence: 5'-CAGUAUGGACACUGUCCAAA-3' (SEQ ID NO: 370).

36. The base editor system of claim 34, wherein the guide comprises or consists of one of the following nucleic acid sequences:

CACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUAA AACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 409) or

CCACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACUA AAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 410).

37. The base editor system of any one of claims 31-35, wherein the guide polynucleotide comprises one or more modified nucleosides at the 5' end and/or the 3' end of the guide.

38. The base editor system of claim 36, wherein the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide.

39. The base editor system of claim 37 wherein the guide polynucleotide comprises two, three, four or more modified nucleosides at the 5' end and/or the 3' end of the guide.

40. The base editor system of claim 36, wherein the guide polynucleotide comprises four modified nucleosides at the 5' end and four modified nucleosides at the 3' end of the guide.

41. The base editor system of any one of claims 36-39, wherein the modified nucleoside comprises a 2'O-methyl or a phosphorothioate.

42. The base editor system of claim 40, wherein the guide polynucleotide comprises or consists essentially of one of the following sequences:

mCsmAsmCsCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUC UACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU

(SEQ ID NO: 409)

or

mCsmCsmAsCCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAU CUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAmUsmUsmUsU

(SEQ ID NO: 410), wherein "m" denotes a 2'-O-methyl and "s" denotes a phosphorothioate.

43. The base editor system of any one of claims 34-41, wherein the guide polynucleotide comprises a nucleic acid sequence, in 5' to 3' orientation, selected from

CCACCAGUAUGGACACUGUC (SEQ ID NO: 371); CACCAGUAUGGACACUGUCC (SEQ ID NO: 372); ACCAGUAUGGACACUGUCCA (SEQ ID NO: 373); CCAGUAUGGACACUGUCCAA (SEQ ID NO: 374); CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370); AGUAUGGACACUGUCCAAAG (SEQ ID NO: 375); GUAUGGACACUGUCCAAAGA (SEQ ID NO: 376); or UAUGGACACUGUCCAAAGAG (SEQ ID NO: 377).

44. The base editor system of any one of claims 34-42, wherein the adenosine deaminase variant domain is internal to the Cas protein.

45. A polynucleotide encoding the adenosine deaminase variant of any one of claims 1-6, the fusion protein or complex of any one of claims 7-30 or the base editor system of any one of claims 31-44.

46. The polynucleotide of claim 45, wherein the polynucleotide comprises one or more modified nucleosides or nucleotides.

47. The polynucleotide of claim 45, wherein the polynucleotide is DNA or RNA.

48. The polynucleotide of claim 46, wherein the polynucleotide comprises a modification selected from the group consisting of 2′-*O*-methyl (2′-OMe), phosphorothioate (PS), 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), and constrained ethyl (S-cEt).

49. A cell comprising the polynucleotide of any one of claims 45-48.

50. A cell comprising the adenosine deaminase variant of any one of claims 1-6, the fusion protein or complex of any one of claims 7-30, or the base editor system of any one of claims 31-44, or the polynucleotide of any one of claims 45-48.

51. The cell of claim 49 or 50, wherein the cell is a hepatocyte, a hepatocyte precursor, or an iPSc-derived hepatocyte.

52. The cell of any one of claims 49-51, wherein the cell expresses a G6PC polypeptide.

53. The cell of any one of claims 49-52, wherein the cell is from a subject having Glycogen Storage Disease Type 1a (GSD1a).

54. The cell of any one of claims 49-53, wherein the cell is a mammalian cell *in vivo*, *ex vivo*, or *in vitro*.

55. The cell of any one of claims 49-54, wherein the cell is a human cell.

56. The cell of claim 49, wherein the fusion protein and the one or more guide polynucleotides form a complex in the cell.

57. A method of treating a genetic disease in a subject in need thereof, the method comprising administering to a cell of the subject the base editor system of any one of claims 31-44 or a polynucleotide encoding the base editor system.

58. A method of treating a genetic disease in a subject in need thereof, the method comprising administering to the subject the cell of any one of claims 49-56.

59. The method of claim 57 or 58, wherein after treatment the cell expresses a G6PC polypeptide capable of catalyzing the hydrolysis of D-glucose 6-phosphate to D-glucose and orthophosphate.

60. The method of claim 59, wherein the cell is autologous, allogeneic, or xenogeneic to the subject, and/or wherein the genetic disease is Glycogen Storage Disease Type 1a (GSD1a).

61. A method for correcting a single nucleotide polymorphism (SNP) in a polynucleotide, the method comprising:

contacting a target nucleotide sequence, at least a portion of which is located in the polynucleotide or its reverse complement, with the base editor system of any one of claims 31-44; and editing the SNP by deaminating the SNP or its complement nucleobase upon targeting of the base editor to the target nucleotide sequence, wherein deaminating the SNP or its complement nucleobase corrects the SNP.

62. The method of claim 61, wherein the SNP is associated with Glycogen Storage Disease Type 1a (GSD1a).

63. The method of claim 60 or 61, wherein the SNP is in the *G6PC* gene.

64. A method of editing a glucose-6-phosphatase (G6PC) polynucleotide comprising a single nucleotide polymorphism (SNP) associated with Glycogen Storage Disease Type 1a (GSD1a), the method comprising contacting the G6PC polynucleotide with a fusion protein or complex of any one of claims 7-29 in a complex with one or more guide polynucleotides, wherein one or more of the guide polynucleotides targets the base editor to effect an A•T to G•C alteration of the SNP associated with GSD1a.

65. The method of any one of claims 61-64, wherein the contacting is in a cell, a eukaryotic cell, a mammalian cell, or human cell.

66. The method of any one of claims 57-64, wherein the SNP changes a glutamine (Q) to a non-glutamine (X) amino acid or changes an arginine (R) to a non-arginine (X) in a G6PC polypeptide.

67. The method of any one of claims 57-64, wherein the SNP results in expression of an G6PC polypeptide having a non-glutamine (X) amino acid at position 347 or a non-arginine (X) amino acid at position 83.

68. The method of any one of claims 57-64, wherein the base editor correction replaces the non-glutamine amino acid (X) at position 347 with a glutamine or the non-arginine amino acid (X) at position 83 with an arginine.

69. The method of any one of claims 57-64, wherein the SNP results in expression of a G6PC polypeptide that prematurely terminates at amino acid position 347 or at a cysteine at position 83.

70. The method of any one of claims 57-64, wherein the SNP encodes one or more of Q347X and/or R83C.

71. The method of any one of claims 57-70, wherein the editing results in less than 0.5% indel formation.

72. The method of any one of claims 57-71, wherein the editing rescues G6PC catalytic activity.

73. The method of claim 64, wherein the guide polynucleotide comprises a nucleic acid sequence, from 5'-3', selected from the group consisting of CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370); CCACCAGUAUGGACACUGUC (SEQ ID NO: 371); CACCAGUAUGGACACUGUCC (SEQ ID NO: 372); ACCAGUAUGGACACUGUCCA (SEQ ID NO: 373); CCAGUAUGGACACUGUCCAA (SEQ ID NO: 374); AGUAUGGACACUGUCCAAAG (SEQ ID NO: 375); GUAUGGACACUGUCCAAAGA (SEQ ID NO: 376); and UAUGGACACUGUCCAAAGAG (SEQ ID NO: 377).

74. The method of any one of claims 57-60, wherein the subject sustains at least a 24 hour fasting period after treatment.

75. A vector comprising the polynucleotide of any one of claims 45-48.

76. The vector of claim 75, wherein the vector is a viral vector.

77. The vector of claim 76, wherein the viral vector is a retroviral vector, adenoviral vector, lentiviral vector, herpesvirus vector, or adeno-associated viral vector (AAV).

78. A composition comprising the fusion protein or complex of any one of claims 7-30, the base editor system of any one of claims 31-44, the polynucleotide of any one of claims 45-48, the cell of any one of claims 49-56, or the vector of any one of claims 75-77.

79. The composition of claim 78, further comprising a pharmaceutically acceptable excipient or carrier.

80. The composition of claim 78 or 79, wherein the one or more guide polynucleotides and the fusion protein are formulated together or separately.

81. The composition of any one of claims 78-80, further comprising a ribonucleoparticle suitable for expression in a mammalian cell.

82. A composition comprising the polynucleotide of any one of claims 45-48.

83. The composition of claim 82, further comprising a pharmaceutically acceptable excipient or carrier.

84. A composition comprising the cell of any one of claims 49-56.

85. The composition of claim 84, further comprising a pharmaceutically acceptable excipient or carrier.

86. The composition of any one of claims 78-85, further comprising a lipid.

87. The composition of claim 86, further wherein the lipid comprises a lipid nanoparticle.

88. A kit comprising the fusion protein or complex of any one of claims 7-30, the base editor system of any one of claims 31-44, the polynucleotide of any one of claims 45-48, the

cell of any one of claims 49-56, the vector of any one of claims 75-77, or the composition of any one of claims 78-87.

89. The kit of claim 88, further comprising written instructions for the use of the kit in the treatment of Glycogen Storage Disease Type 1a (GSD1a).

90. The fusion protein or complex of any one of claims 7-30, the base editor system of any one of claims 31-44, the polynucleotide of any one of claims 45-48, the cell of any one of claims 49-56, the vector of any one of claims 75-77, or the composition of any one of claims 78-85, wherein the base editor comprises an mRNA sequence as set forth in SEQ ID NO: 396.

91. The fusion protein or complex of any one of claims 7-30, the base editor system of any one of claims 31-44, the polynucleotide of any one of claims 45-48, the cell of any one of claims 49-56, the vector of any one of claims 75-77, or the composition of any one of claims 78-85, wherein the base editor comprises a DNA sequence as set forth in SEQ ID NO: 397.

92. The fusion protein or complex of any one of claims 7-30, the base editor system of any one of claims 31-44, the polynucleotide of any one of claims 45-48, the cell of any one of claims 49-56, the vector of any one of claims 75-77, or the composition of any one of claims 78-85, wherein the base editor comprises an amino acid sequence as set forth in SEQ ID NO: 398.

93. A modified guide RNA (gRNA) comprising modified nucleotides, wherein the guide comprises from 5' to 3' a polynucleotide sequence selected from the group consisting of:

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACU AAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 404);

UUUCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCU ACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 405);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAG GCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 406);

CCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUAC UAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 407);

ACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 408);

CACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCU ACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 409);

CCACCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUC UACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 410);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAGAAAUACAGAAUCUACUAAAA CAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 411);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGCGGAAACGCAGAAUCUACUAAAA CAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 412);

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCGAAAGAAUCUACUAAAACAAGGCAA AAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 413);

CAGUAUGGACACUGUCCAAAGUUUUAGUACCCGAAAGCAUCUACUAAAACAAGGCAA AAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 414); and

CAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAAUUACAGAAUCUACU AAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUUUU (SEQ ID NO: 415).

94. The modified gRNA of claim 93, wherein the guide comprises at least about 50%-75% modified nucleotides.

95. The modified gRNA of claim 93, wherein the guide comprises at least about 85% or more modified nucleotides.

96. The modified gRNA of claim 93, wherein at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified.

97. The modified gRNA of claim 96, wherein at least about 3-5 contiguous nucleotides at each of the 5' and 3' termini of the gRNA are modified.

98. The modified gRNA of claim 93, wherein at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified.

99. The modified gRNA of claim 93, wherein at least about 50% of the nucleotides present in a direct repeat or anti-direct repeat are modified.

100. The modified gRNA of claim 93, wherein at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified.

101. The modified gRNA of claim 93, wherein at least about 100 of the nucleotides present in a direct repeat or anti-direct repeat are modified.

102. The modified gRNA of claim 93, wherein at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified.

103. The modified gRNA of claim 93, wherein at least about 50% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified.

104. The modified gRNA of claim 93, wherein the guide comprises a variable length protospacer.

105. The modified gRNA of claim 93, wherein the guide comprises a 20-40 nucleotide protospacer.

106. The modified gRNA of claim 93, wherein the guide comprises a protospacer comprising at least about 20-25 nucleotides or at least about 30-35 nucleotides.

107. The modified gRNA of claim 93, wherein the protospacer comprises modified nucleotides.

108. The modified gRNA of claim 93, wherein the guide comprises two or more of the following:

at least about 1-5 nucleotides at the 5' end of the gRNA are modified and at least about 1-5 nucleotides at the 3' end of the gRNA are modified;

at least about 20% of the nucleotides present in a direct repeat or anti-direct repeat are modified;

at least about 50-75% of the nucleotides present in a direct repeat or anti-direct repeat are modified;

at least about 20% or more of the nucleotides present in a hairpin present in the gRNA scaffold are modified;

a variable length protospacer; and

a protospacer comprising modified nucleotides.

109. The modified gRNA of any one of claims 93-108, wherein the gRNA comprises one or more modifications selected from the group consisting of 2′-*O*-methyl (2′-OMe), phosphorothioate (PS), 2′-*O*-methyl thioPACE (MSP), 2′-*O*-methyl-PACE (MP), 2′-fluoro RNA (2′-F-RNA), and constrained ethyl (S-cEt).

110. The modified gRNA of claim 109, wherein the gRNA comprises 2′-*O*-methyl or phosphorothioate modifications.

111. The modified gRNA of claim 109, wherein the gRNA comprises 2′-*O*-methyl and phosphorothioate modifications.

112. The modified gRNA of claim 109, wherein the modifications increase base editing by at least about 2 fold.

113. A modified guide RNA (gRNA) comprising a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUmAmGmUACUCmUGmUmAmAmUGmAAAmAmUmUmACmAGAAUCUACmUmAAAACAAGGCAAmAAUGmCCmGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404);

mUsmUsmUsCAGmUAUmGmGmACAmCUGUCCAAAmGUUUUmAmGmUACUCmUGmUmAmAmUGmAAAmAmUmUmACmAGAAUCUACmUmAAAACAAGGCAAmAAUGmCCmGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 405);

mCsmAsGsmUAUmGmGmAmCAmCmUGUCCAAmAmGUmUUmUmAmGmUACUmCmUmGmUmAmAmUGmAmAmAmAmUmUmACmAmGAAmUCUACmUmAmAAACAAGmGCAAmAAUGmCmCmGUGmUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsGsmUmAUmGmGmAmCAmCUGUCCAAmAmGUUUmUAmGmUACUmCmUmGmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAAmUCUACUmAmAAACAAmGmGmCmAmAmAAUmGmCmCGUGmUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 404); or

mCsmAsGsmUmAUmGmGmAmCAmCmUGUCmCmAAmAmGmUmUmUmUmAmGmUAmCmUmCmUmGmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAmAm

UCUACmUmAmAmAAmCAmAmGmGmCmAmAmAAUmGmCmCmGmUGmUmUmU mAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsm UsmUsmU (SEQ ID NO: 404), wherein "m" denotes a 2'-O-methyl and "s" denotes a phosphorothioate.

114. A modified guide RNA (gRNA) comprising a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGmUmAm AmUmGmAmAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGC CGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGmUmAm AmUmGmAmAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGC CGUGUmUmUmAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGm AmGmAmUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUAGUACmUmCmU mGmUmAmAmUmGmAmAmAmAmUmUmAmCmAmGmAAUCUACUAAAACAAGGC AAAAUGCCmGUGUmUmUAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmG mGmCmGmAmGmAUsmUsmUsmU (SEQ ID NO: 404);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACmUmCmUmGGmAmA mAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCGUGUmUmUmAmUmCmU mCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAmUsmUsmUsmU (SEQ ID NO: 406); or

mCsmAsmGsmUAUmGmGmACAmCUGUCCAAAmGUUUUAGUACmUmCmU mGmGmAmAmAmCmAmGmAAUCUACUAAAACAAGGCAAAAUGCCmGUGUmUm UAmUmCmUmCmGmUmCmAmAmCmUmUmGmUmUmGmGmCmGmAmGmAUsmUs mUsmU (SEQ ID NO: 406), wherein "m" denotes a 2'-O-methyl and "s" denotes a phosphorothioate.

115. A modified guide RNA (gRNA) comprising a nucleic acid sequence, from 5' to 3', selected from

mCsmCsmAsGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAA AUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUU GUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 407);

mAsmCsmCsAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAA AAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACU UGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 408);

mCsmAsmCsCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGA AAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAAC UUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 409); or

mCsmCsmAsCCAGUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUG AAAAUUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAA CUUGUUGGCGAGAmUsmUsmUsU (SEQ ID NO: 410), wherein "m" denotes a 2'-O-methyl and "s" denotes a phosphorothioate.

116. A modified guide RNA (gRNA) comprising a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAGAAAUAC AGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGG CGAGAUsmUsmUsmU (SEQ ID NO: 411);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGCG GAAA CGCAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGU UGGCGAGAUsmUsmUsmU (SEQ ID NO: 412);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGGAAACAGAA UCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAG AUsmUsmUsmU (SEQ ID NO: 406);

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCGAAAGAAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmU smUsmU (SEQ ID NO: 413); or

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACCCGAAAGCAUCUA CUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUGUUGGCGAGAUsmU smUsmU_(SEQ ID NO: 414), wherein "m" denotes a 2'-O-methyl and "s" denotes a phosphorothioate.

117. A modified guide RNA (gRNA) comprising a nucleic acid sequence, from 5' to 3', selected from

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAA UUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUG UUGGCGAGAUsmUsmUsmU (SEQ ID NO: 415); or

mCsmAsmGsUAUGGACACUGUCCAAAGUUUUAGUACUCUGUAAUGAAAA UUACAGAAUCUACUAAAACAAGGCAAAAUGCCGUGUUUAUCUCGUCAACUUG UUGGCGAGAmUsmUsmUsU (SEQ ID NO: 415), wherein "m" denotes a 2'-O-methyl and "s" denotes a phosphorothioate.

118. A formulation comprising a lipid nanoparticle comprising an mRNA expressing a base editor and a gRNA, wherein the base editor comprises a Cas9 domain and at least one adenosine deaminase variant comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N with reference to SEQ ID NO: 1: MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase; and the gRNA comprises CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370).

119. A formulation comprising a lipid nanoparticle comprising an mRNA expressing a base editor, wherein the base editor comprises a Cas9 domain and at least one adenosine deaminase variant comprising V82G, Y147T/D, Q154S, and one or more of L36H, I76Y, F149Y, N157K, and D167N with reference to SEQ ID NO: 1: MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA LRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 1), or corresponding alterations in another adenosine deaminase; and a gRNA comprising CAGUAUGGACACUGUCCAAA (SEQ ID NO: 370).

120. The formulation of claim 118 or 119, wherein the adenosine deaminase variant domain comprises the following combination of alterations I76Y + V82G + Y147D + F149Y + Q154S + D167N of SEQ ID NO: 1, or corresponding alterations in another adenosine deaminase.

121. The formulation of any one of claims 118-120, wherein the gRNA comprises 2'-O-methyl and/or phosphorothioate modifications.

122. The formulation of claim 121, wherein the gRNA comprises 2′-*O*-methyl and phosphorothioate modifications.

123. The formulation of claim 122, wherein the mRNA comprises one or more pseudouridines.

124. The formulation of claim 122, wherein the mRNA comprises an N1-methylpseudouridine (m1Ψ).

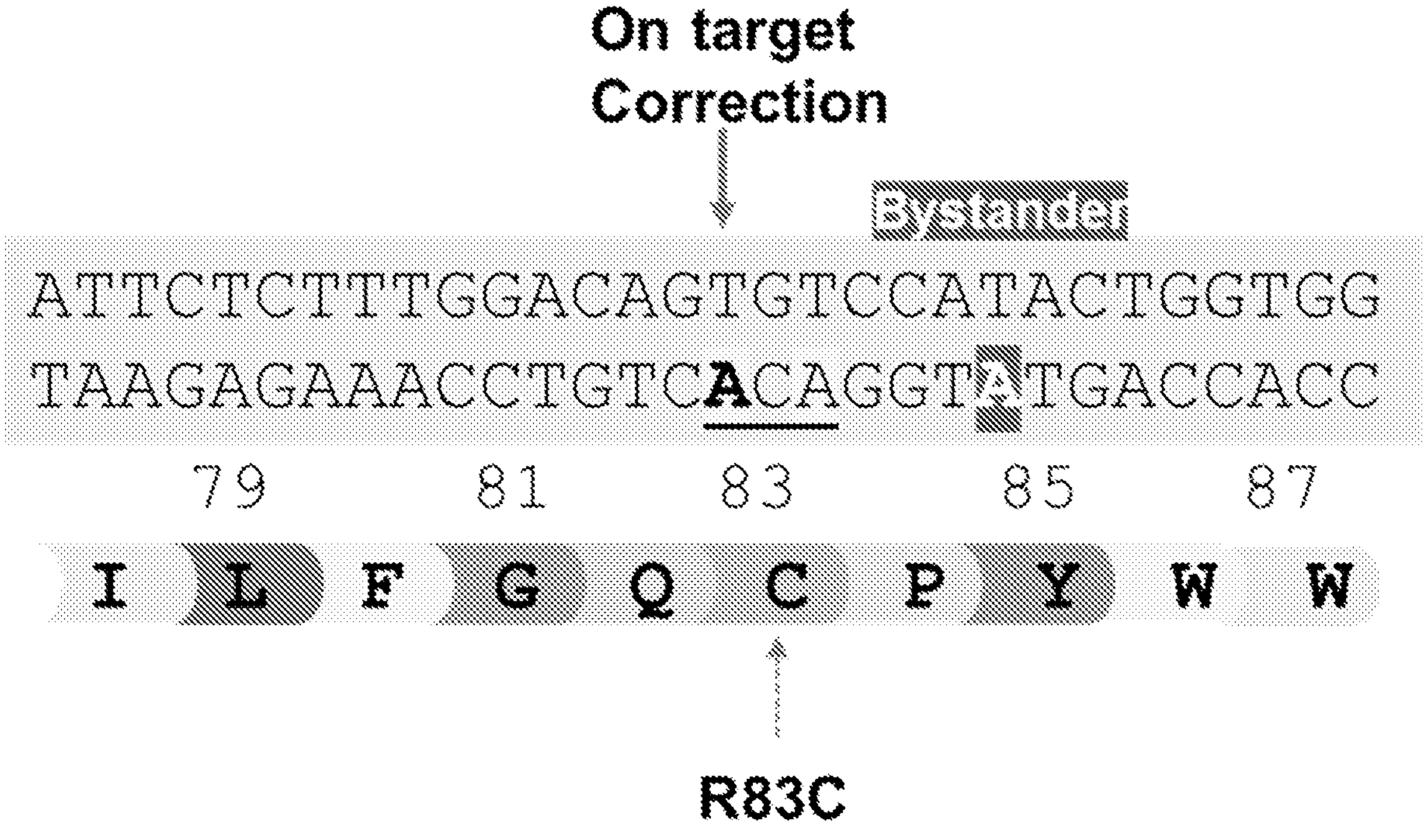

On target
Correction
Silent
ATTCTCTTTGGACAGTGTCCATACTGGTGG
TAAGAGAAACCTGTCACAGGTATGACCACC
79 81 83 85 87
I L F G Q C P Y W W
R83C

FIG. 1

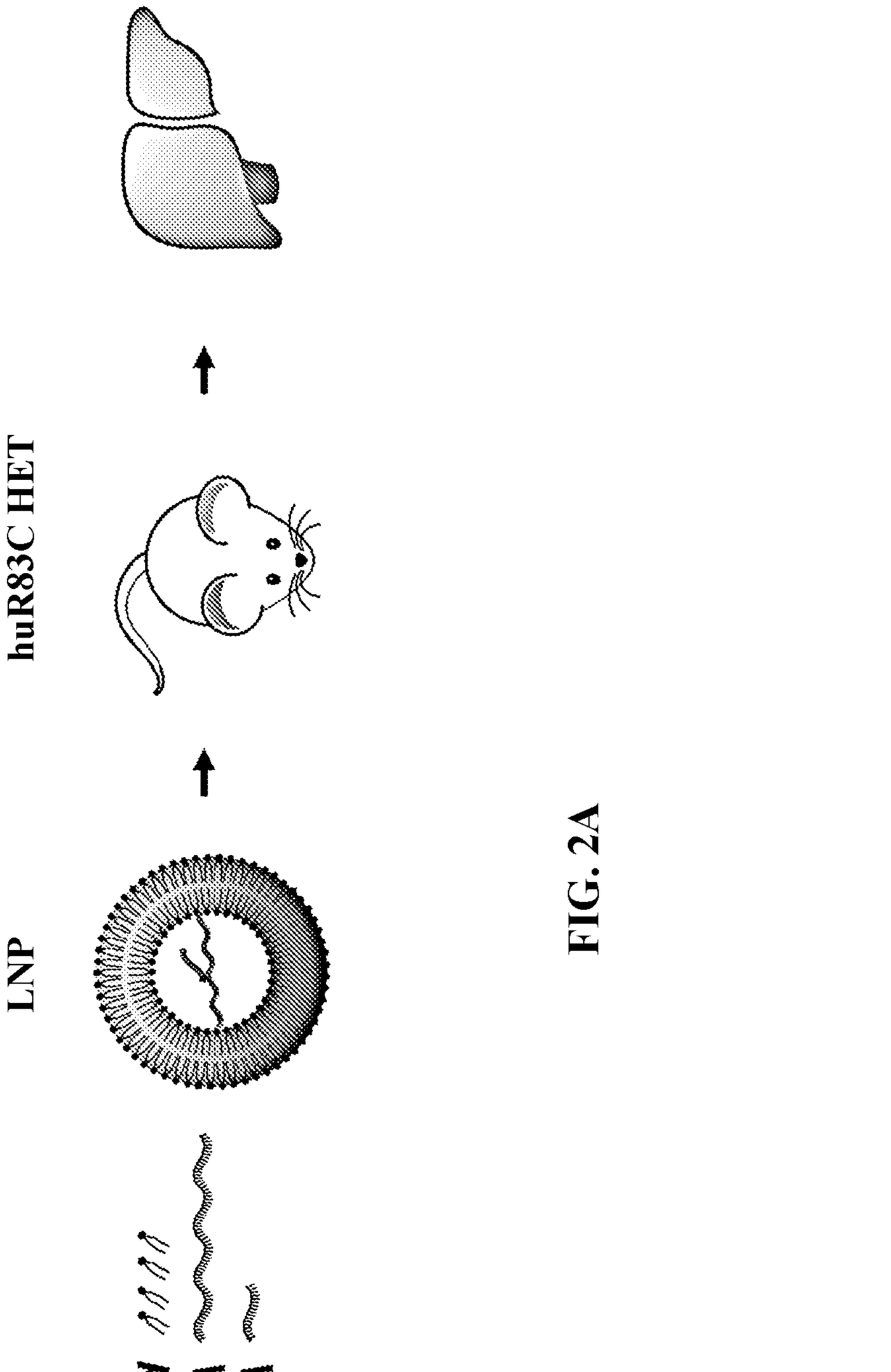


FIG. 2A

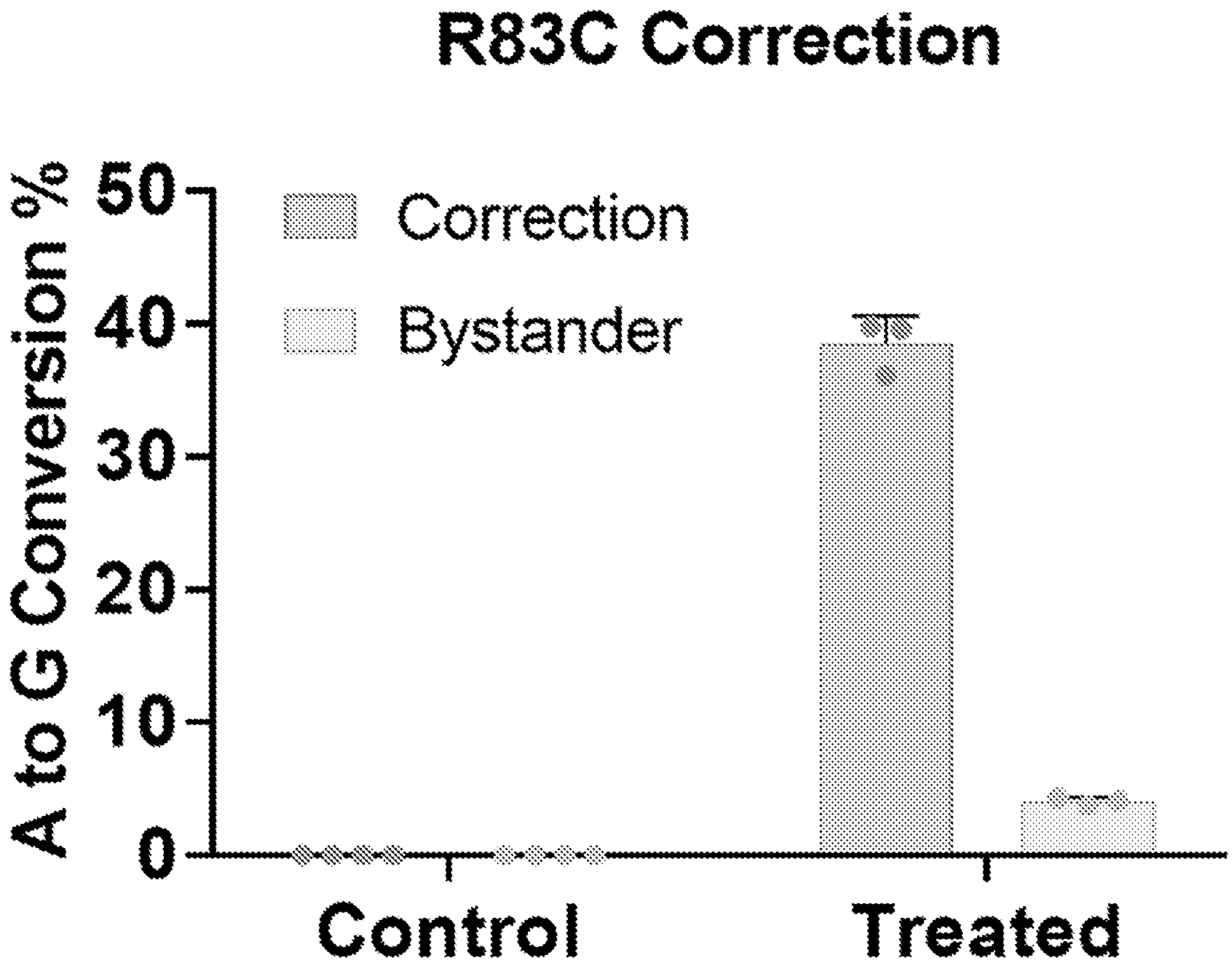

R83C Correction
A to G Conversion %
0
10
20
30
40
50
Correction
Bystander
Control
Treated

FIG. 2B

R83C correction in a transgenic mouse model for GSD1a

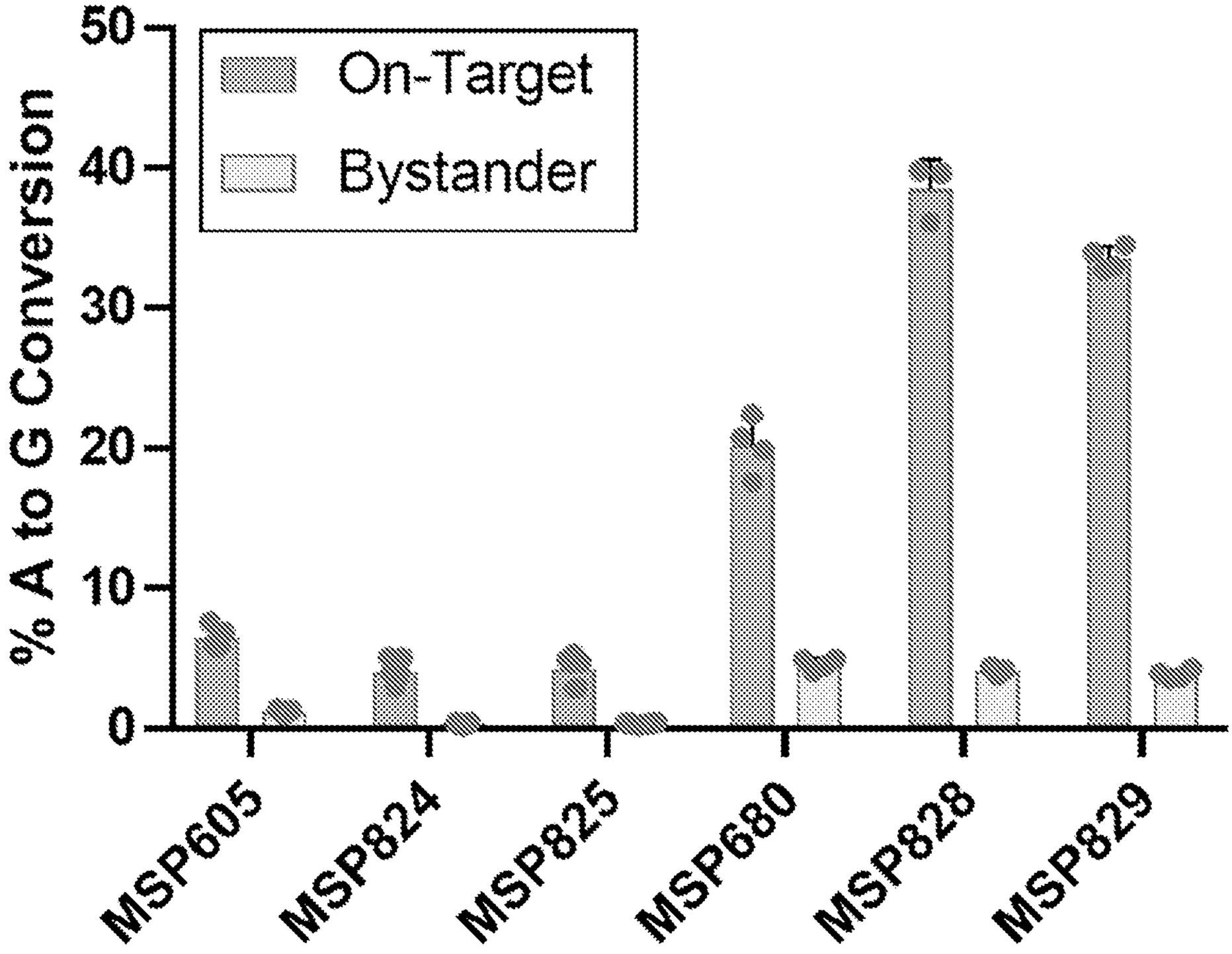


FIG. 3

FIG. 4

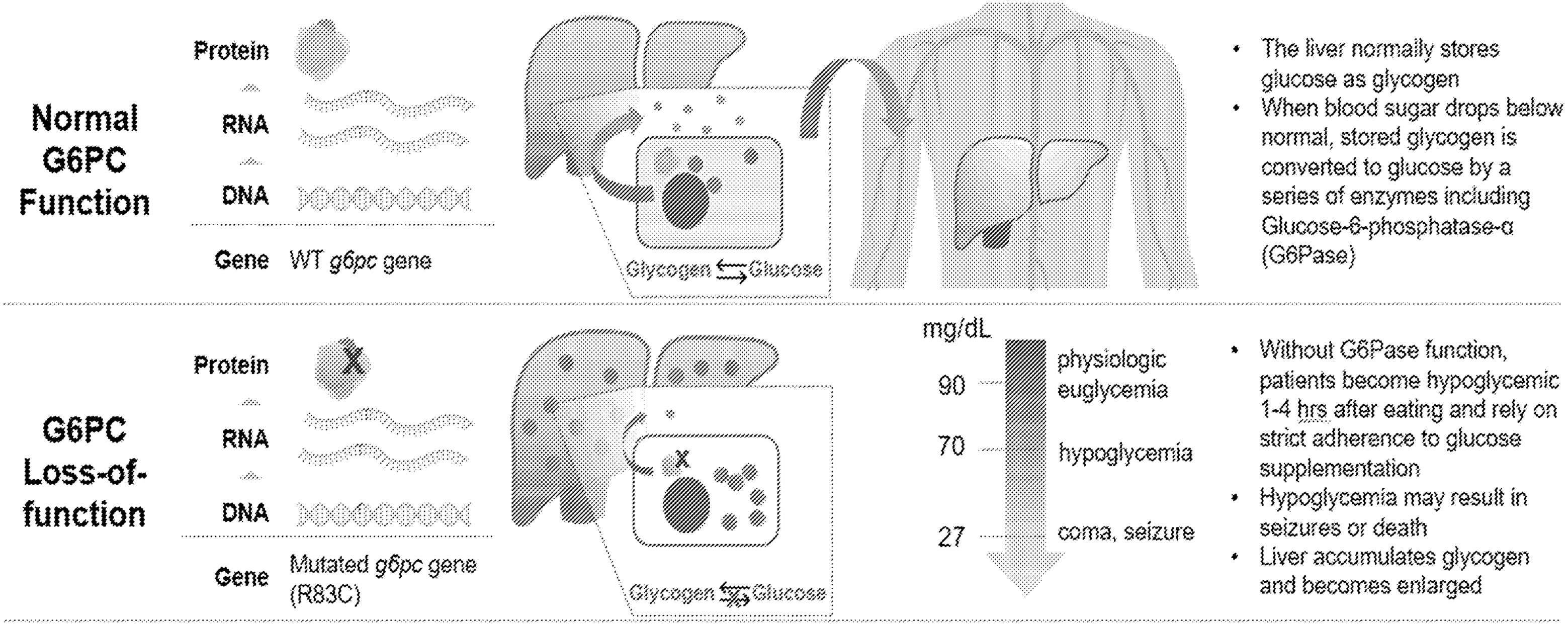

Molecular
Liver
Circulation
Outcome
Normal G6PC Function
Protein
RNA
DNA
Gene
WT g6pc gene
Glycogen ⇆ Glucose
• The liver normally stores glucose as glycogen
• When blood sugar drops below normal, stored glycogen is converted to glucose by a series of enzymes including Glucose-6-phosphatase-α (G6Pase)
G6PC Loss-of-function
Protein
RNA
DNA
Gene
Mutated g6pc gene (R83C)
Glycogen ⇆ Glucose
mg/dL
90
70
27
physiologic euglycemia
hypoglycemia
coma, seizure
• Without G6Pase function, patients become hypoglycemic 1-4 hrs after eating and rely on strict adherence to glucose supplementation
• Hypoglycemia may result in seizures or death
• Liver accumulates glycogen and becomes enlarged

FIG. 5

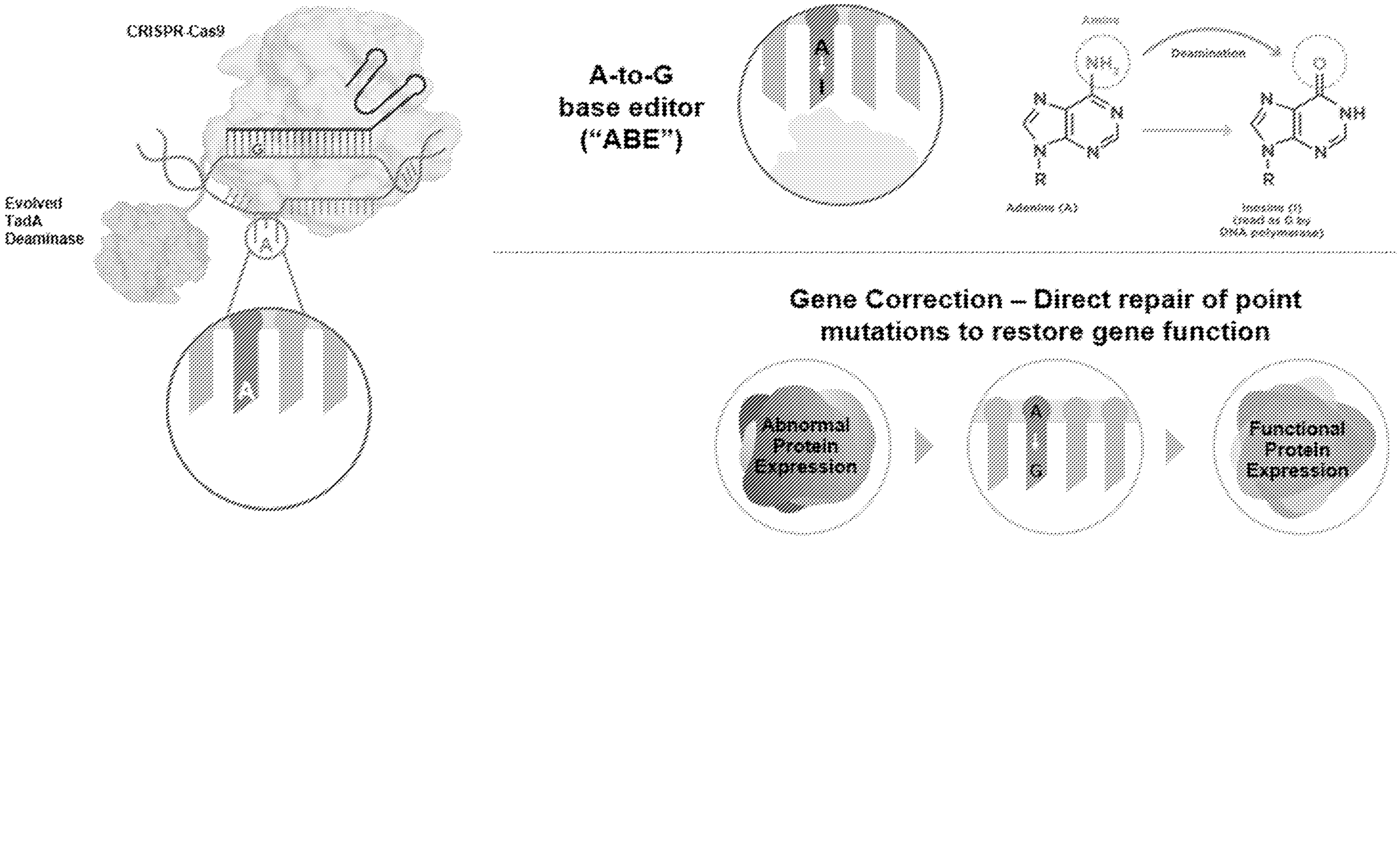
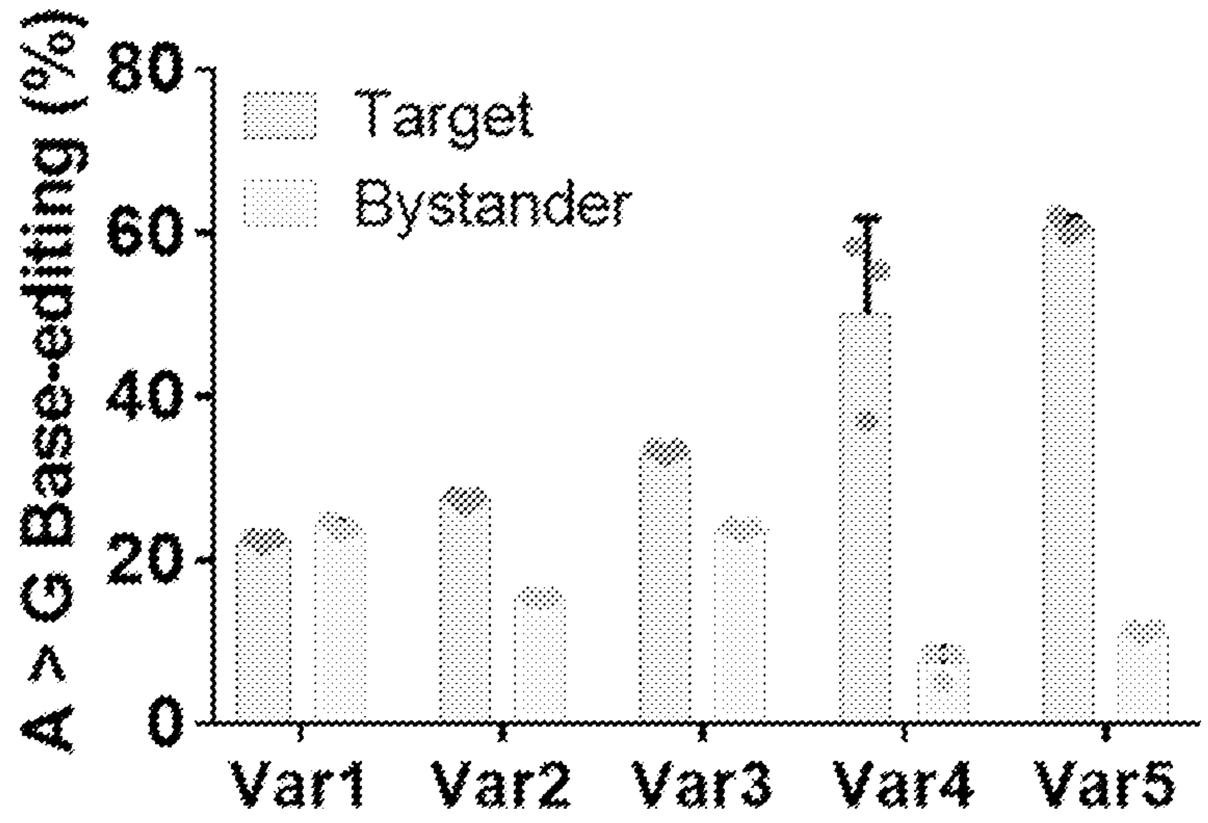

Base editor binds the target DNA and exposes an editing window
CRISPR-Cas9
Evolved TadA Deaminase
Deaminase chemically modifies target base, A>G edit can be made permanent by DNA repair/replication
A-to-G base editor ("ABE")
A
I
Deamination
Gene Correction – Direct repair of point mutations to restore gene function
Abnormal Protein Expression
A
G
Functional Protein Expression

FIG. 6A

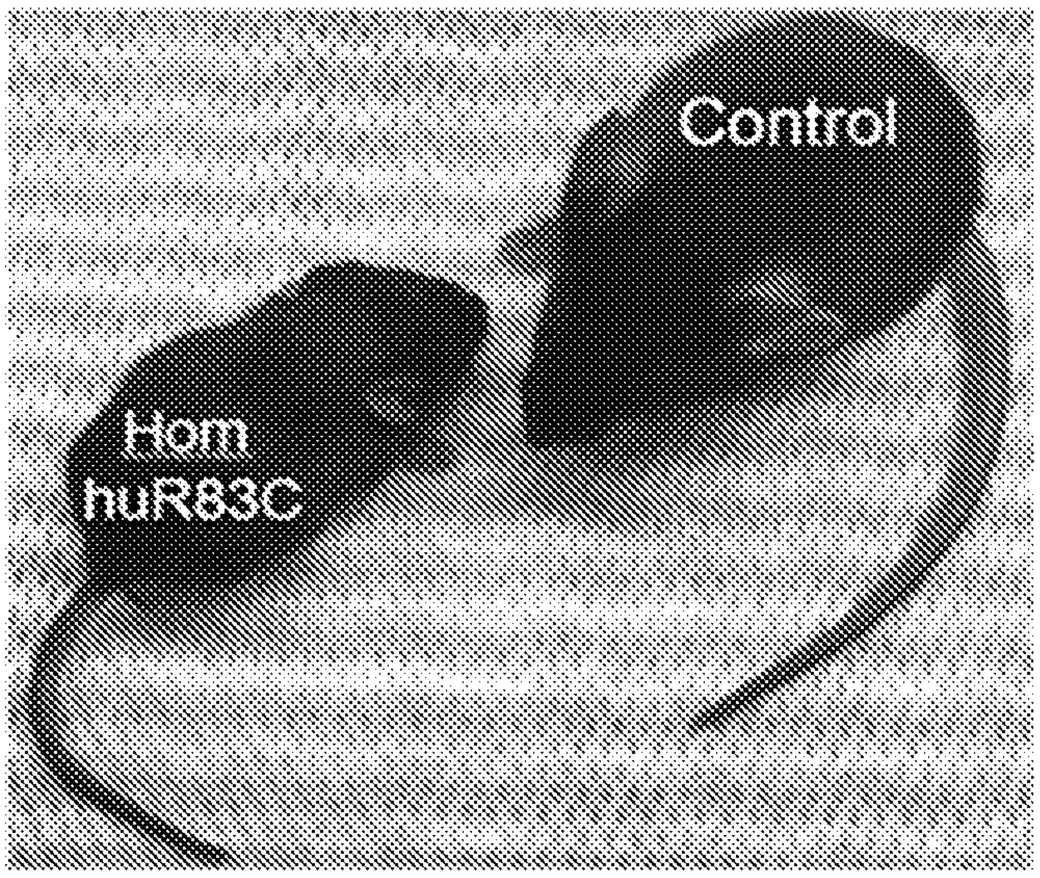


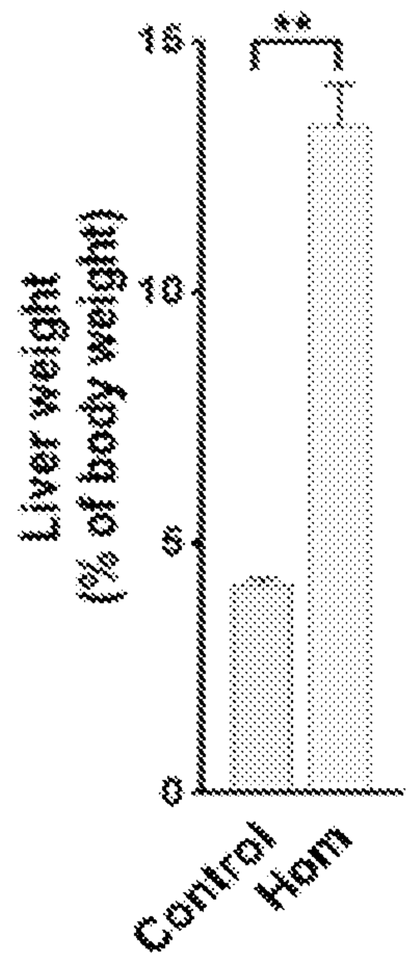
FIG. 6B

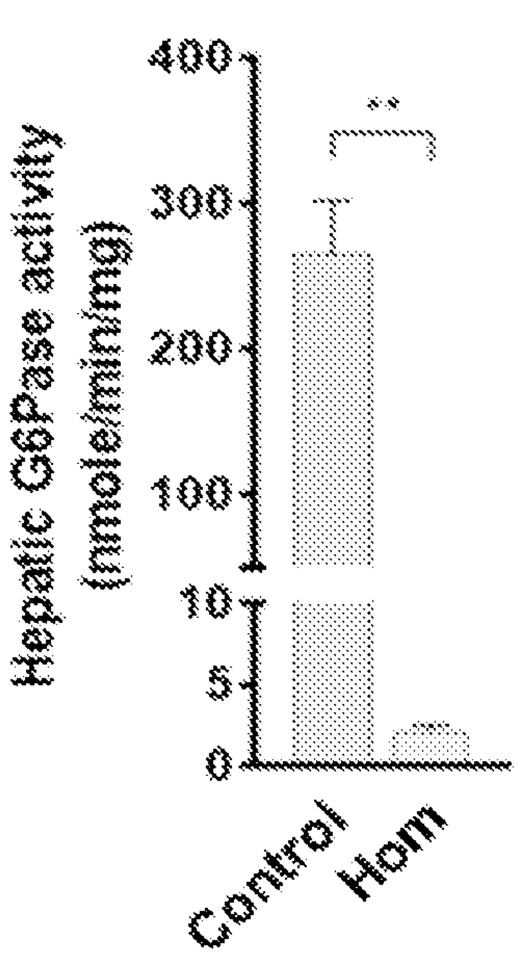


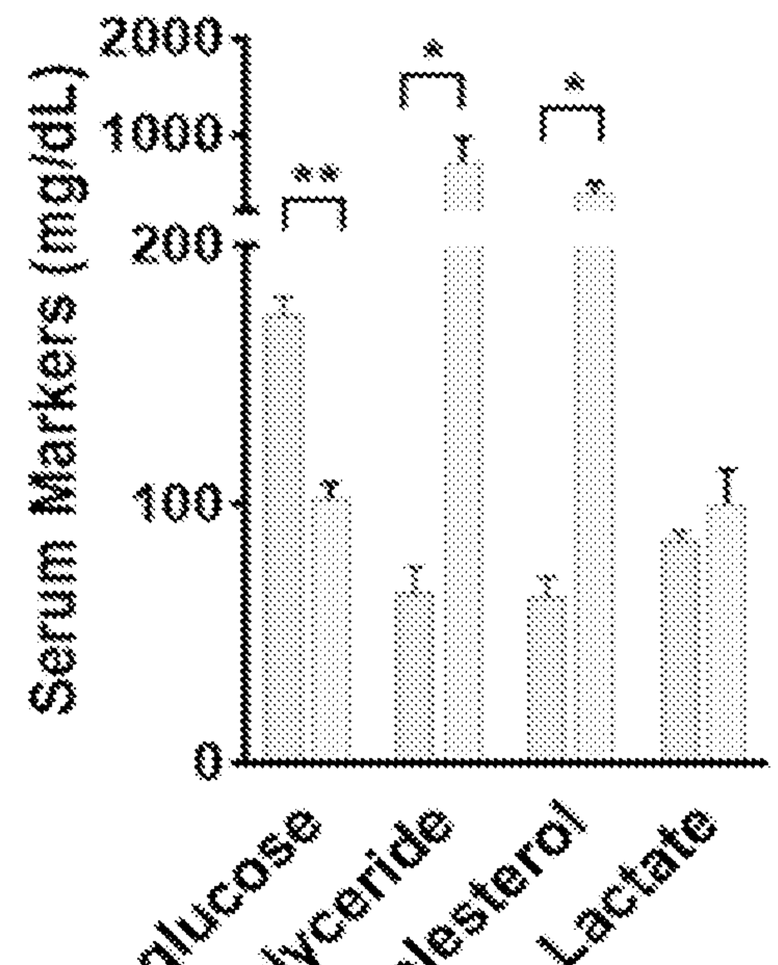
- Lead optimization yields **~60% targeted base-editing efficiency, reduced bystander editing**

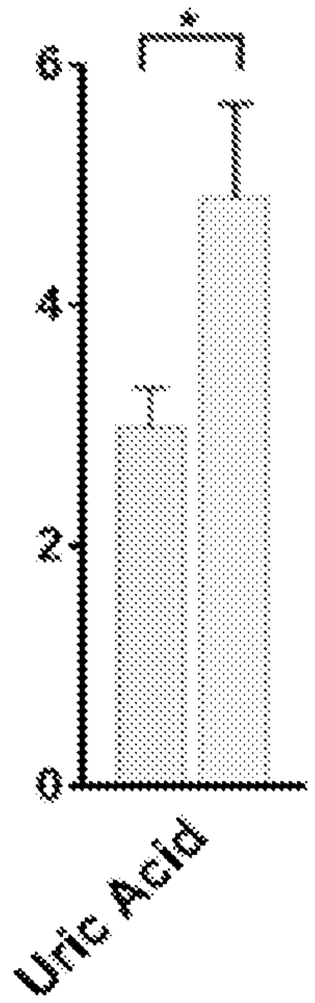
FIG. 7


Control
Hom
huR83C


Liver weight
(% of body weight)
0
5
10
15
**
Control
Hom


Hepatic G6Pase activity
(nmole/min/mg)
0
5
10
100
200
300
400
**
Control
Hom


Serum Markers (mg/dL)
0
100
200
1000
2000
**
*
*
Blood glucose
Triglyceride
Cholesterol
Lactate


0
2
4
6
*
Uric Acid

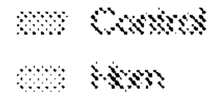

Control
Hom

FIG. 8A

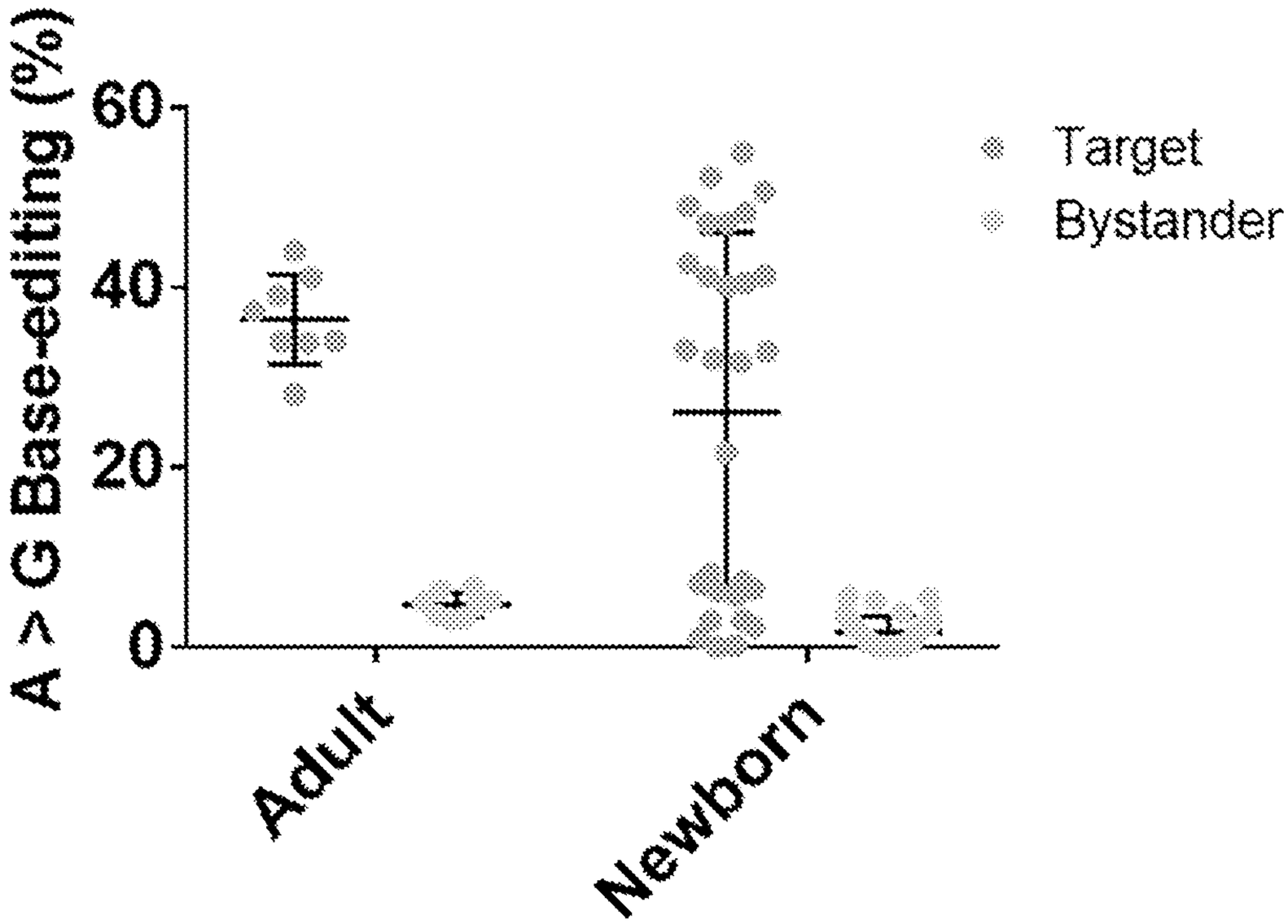

A > G Base-editing (%)
60
40
20
0
Adult
Newborn
Target
Bystander

FIG. 8B

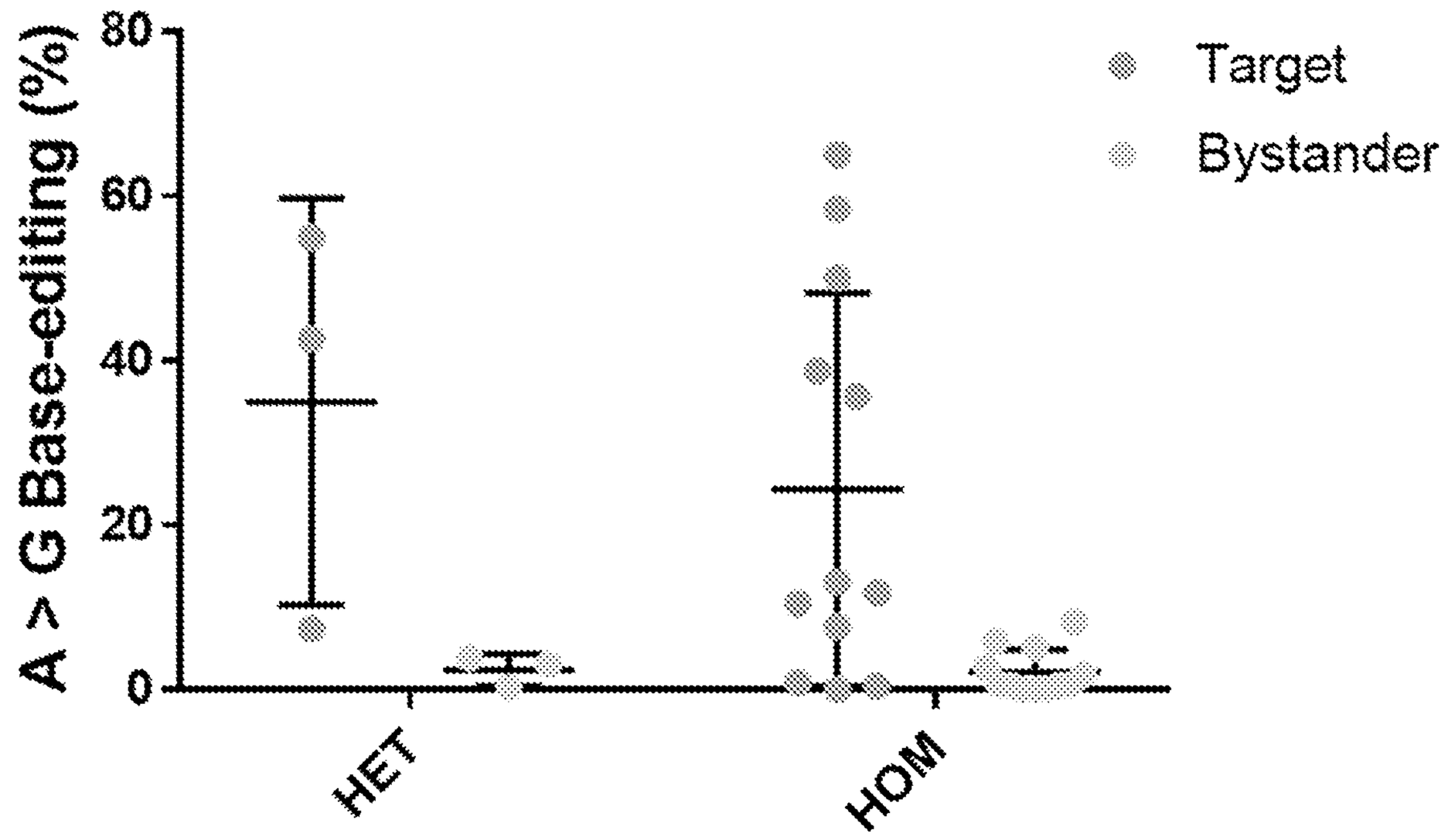

A > G Base-editing (%)
80
60
40
20
0
HET
HOM
Target
Bystander

FIG. 9A

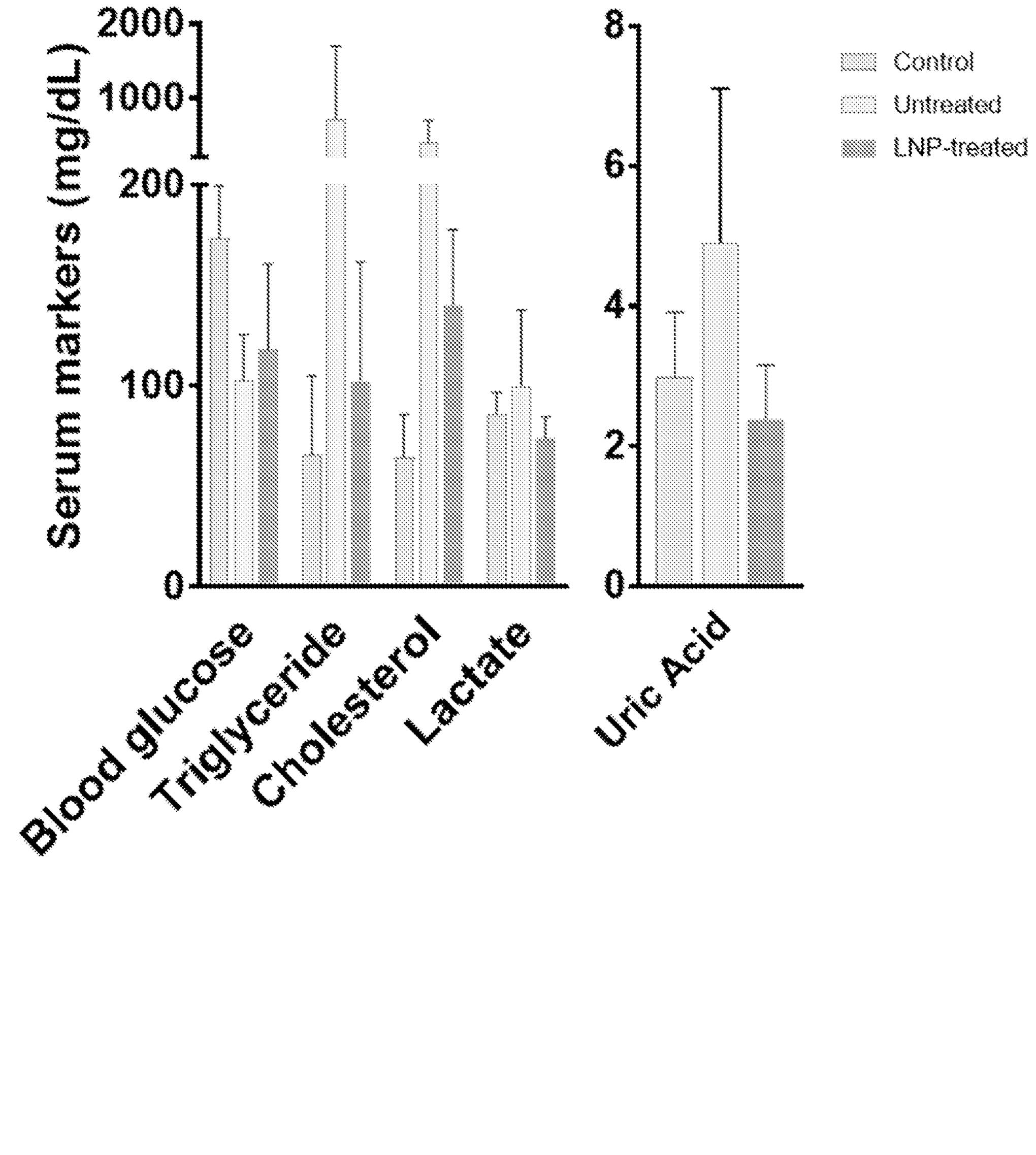

Serum markers (mg/dL)
2000
1000
200
100
0
8
6
4
2
0
Control
Untreated
LNP-treated
Blood glucose
Triglyceride
Cholesterol
Lactate
Uric Acid

FIG. 9B

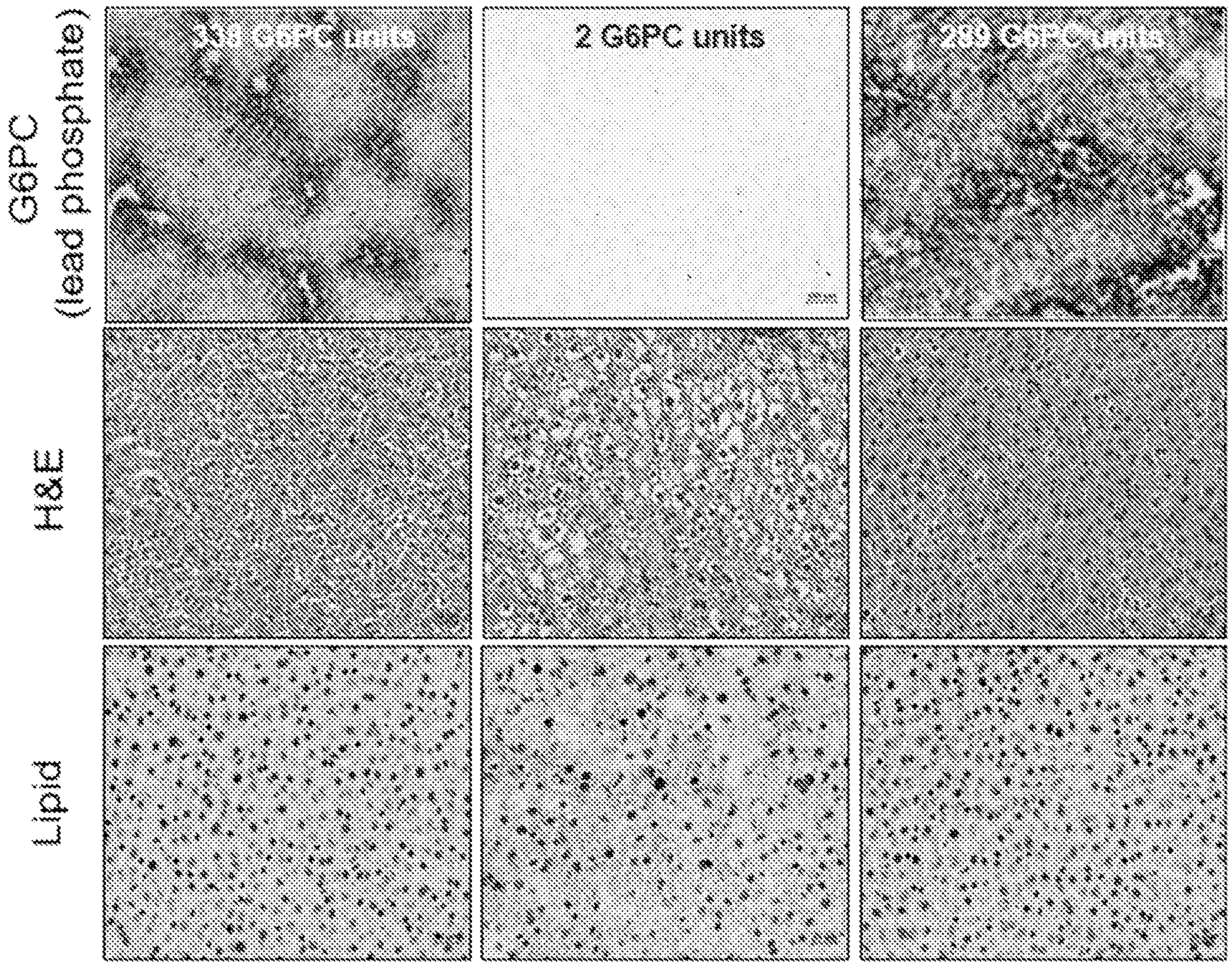
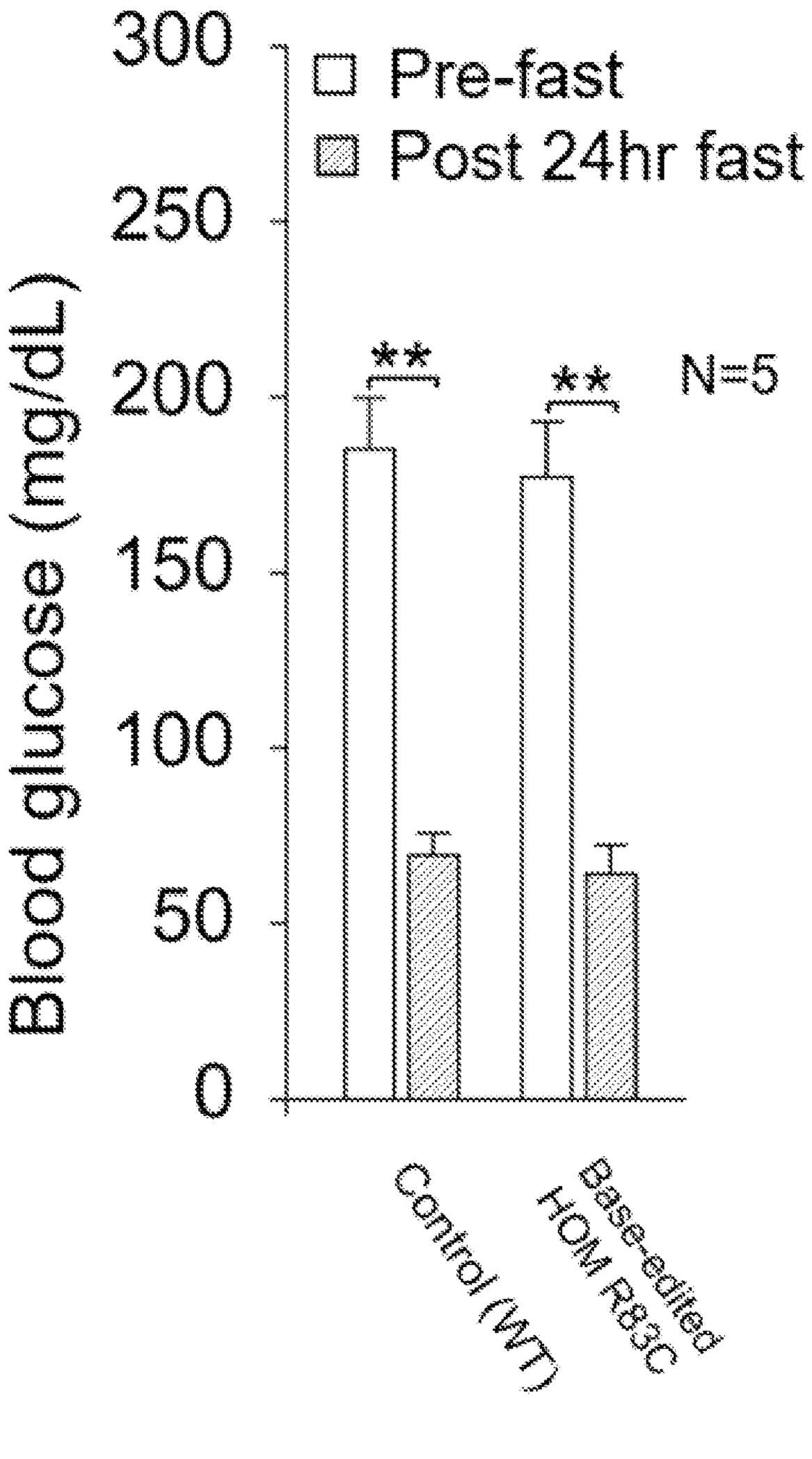


FIG. 10


300
Pre-fast
Post 24hr fast
Blood glucose (mg/dL)
250
200
150
100
50
0
**
**
N=5
Control (WT)
Base-edited
HOM R83C

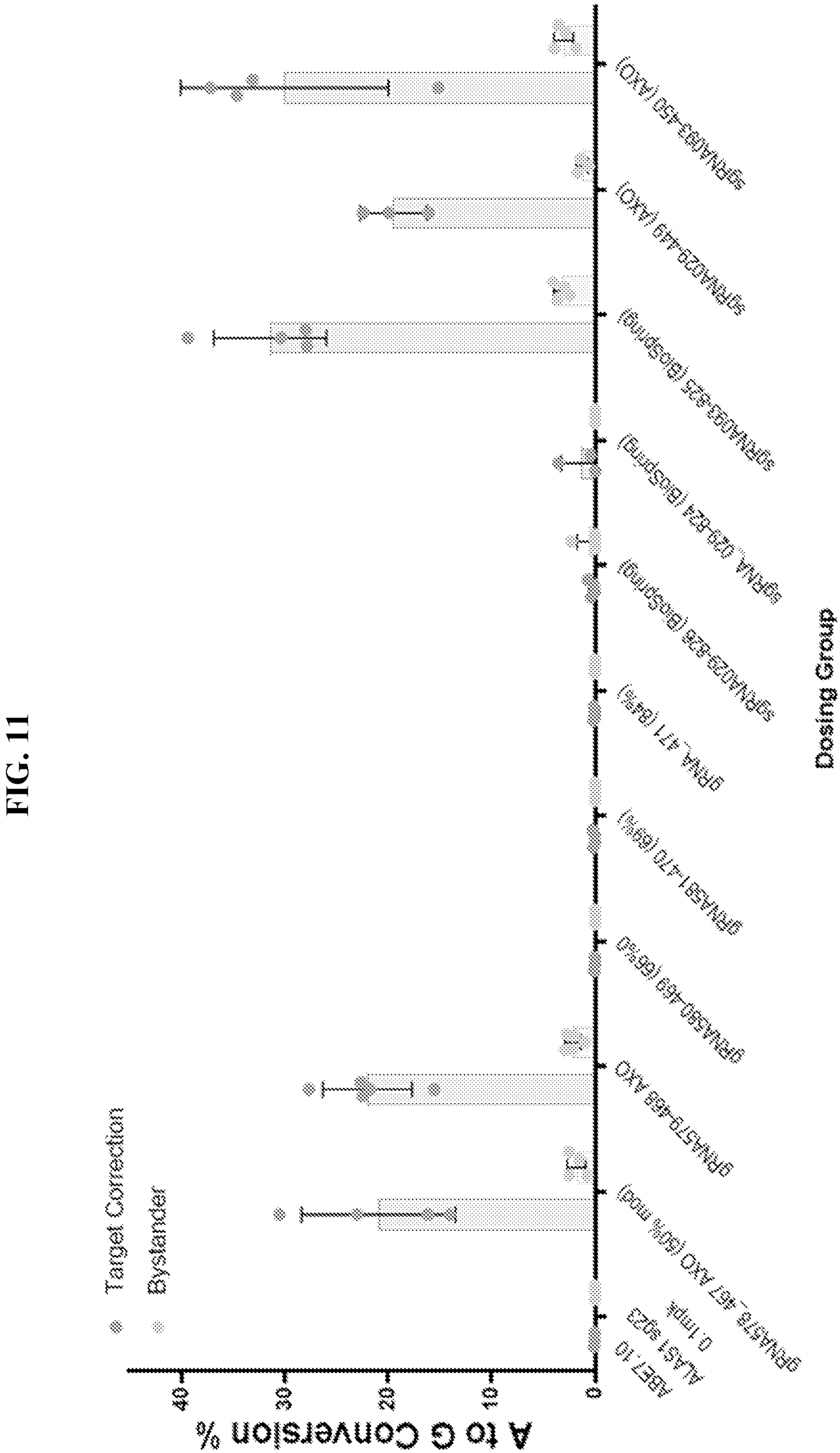
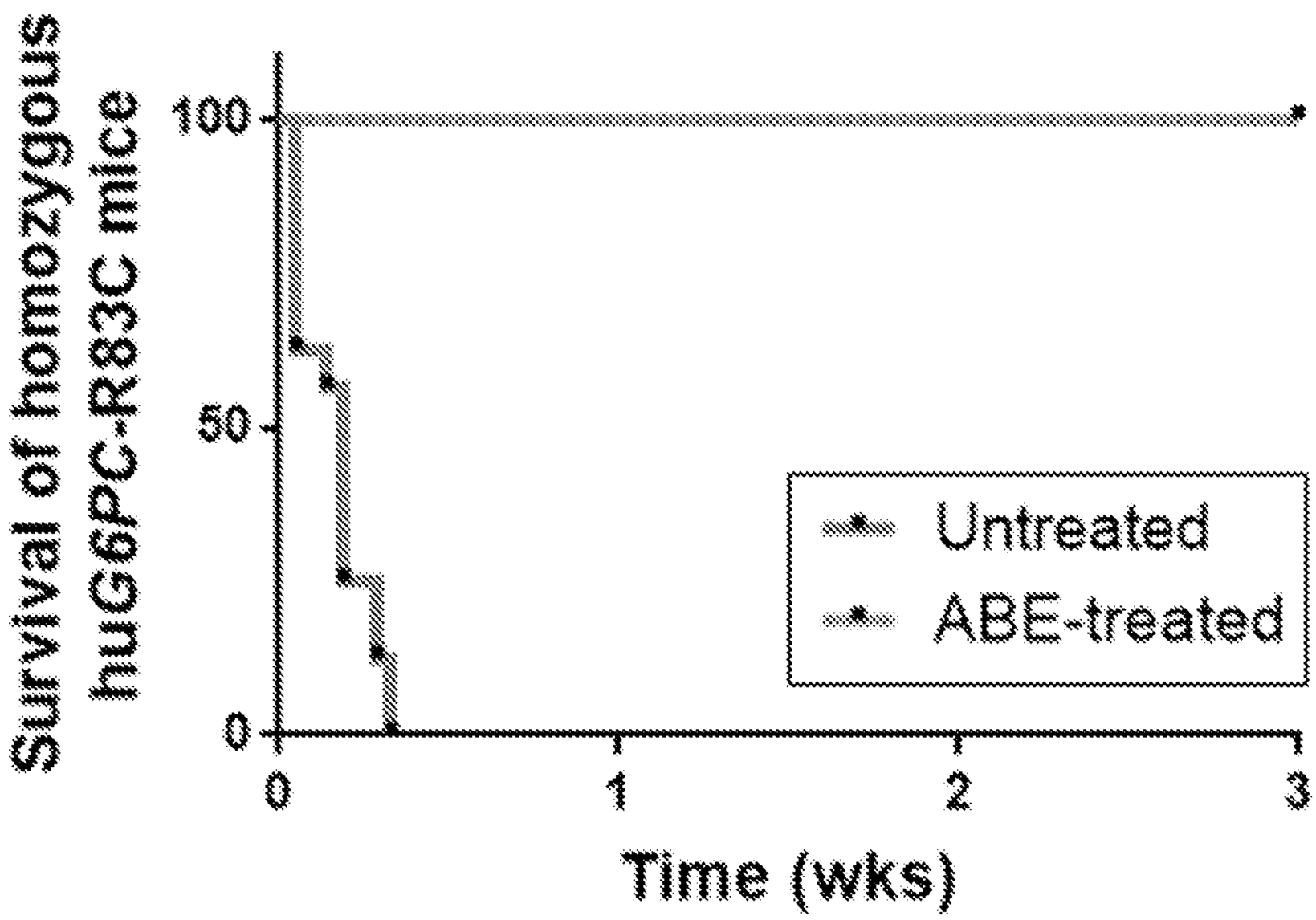

A to G Conversion %
0
10
20
30
40
Target Correction
Bystander
Dosing Group
AAV-ABE8e-SpCas9-NG (1e13 vg/kg)
AAV-ABE8e-SpCas9-NG (5e13 vg/kg)
AAV-ABE8e-SpCas9-NG (1e14 vg/kg)
AAV-ABE8e-SpCas9-NG (5e14 vg/kg)
AAV-ABE8e-SpCas9-NG (1e15 vg/kg)
AAV-ABE8e-SpCas9-NG (5e15 vg/kg)
AAV-ABE8e-SpCas9-NG (1e16 vg/kg)
AAV-ABE8e-SpCas9-NG (5e16 vg/kg)
Untreated

FIG. 11

FIG. 12


Kaplan-Meier Survival Estimates
Survival of homozygous huG6PC-R83C mice
100
50
0
0
1
2
3
Time (wks)
Untreated
ABE-treated

SEQUENCE LISTING

<110> BEAM THERAPEUTICS INC.

<120> COMPOSITIONS AND METHODS FOR TREATING GLYCOGEN STORAGE DISEASE TYPE 1A

<130> 180802.047002/PCT

<150> 63/091,891
<151> 2020-10-14

<150> 63/248,081
<151> 2021-09-24

<160> 427

<170> PatentIn version 3.5

<210> 1
<211> 167
<212> PRT
<213> Escherichia coli

<220>
<221> misc_feature
<222> (1)..(167)
<223> TadA*7.10 (ecTadA) polypeptide sequence

<400> 1

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 Ala Lys Thr Gly
            100                 105                 110


Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160


Lys Ala Gln Ser Ser Thr Asp
                165


<210>  2
<211>  4092
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20
       at P153 polynucleotide sequence


<220>
<221>  misc_feature
<222>  (1)..(6)
<223>  Kozak sequence

<220>
<221>  misc_feature
<222>  (7)..(57)
<223>  Sequence encoding N-terminal nuclear localization signal

<220>
<221>  misc_feature
<222>  (58)..(513)
<223>  Sequence encoding BhCas12b

<220>
<221>  misc_feature
<222>  (514)..(531)
<223>  Sequence encoding a linker

<220>
<221>  misc_feature
```

<222> (532)..(1029)
<223> Sequence encoding TadA*8

<220>
<221> misc_feature
<222> (1030)..(1089)
<223> Sequence encoding Xten20 linker

<220>
<221> misc_feature
<222> (1090)..(3954)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (3955)..(4002)
<223> Sequence encoding C-terminal NLS

<220>
<221> misc_feature
<222> (4003)..(4008)
<223> Seqquence encoding GS linker

<220>
<221> misc_feature
<222> (4009)..(4089)
<223> Sequence encoding 3x hemagglutinin (HA) tag

<400> 2
gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat cccggaggct ctggaggaag ctccgaagtc 540

gagttttccc atgagtactg gatgagacac gcattgactc tcgcaaagag ggctcgagat 600

gaacgcgagg tgcccgtggg ggcagtactc gtgctcaaca atcgcgtaat cggcgaaggt 660

tggaataggg caatcggact ccacgacccc actgcacatg cggaaatcat ggcccttcga 720

cagggagggc ttgtgatgca gaattatcga ctttatgatg cgacgctgta cgtcacgttt 780

gaaccttgcg taatgtgcgc gggagctatg attcactccc gcattggacg agttgtattc 840

ggtgttcgca acgccaagac gggtgccgca ggttcactga tggacgtgct gcatcatcca 900

ggcatgaacc accgggtaga aatcacagaa ggcatattgg cggacgaatg tgcggcgctg 960

ttgtgtcgtt tttttcgcat gcccaggcgg gtctttaacg cccagaaaaa agcacaatcc 1020

tctactgacg gctcttctgg atctgaaaca cctggcacaa gcgagagcgc caccctgag 1080

agctctggct cctgggaaga agagaagaag aagtgggaag aagataagaa aaaggacccg 1140

ctggccaaga tcctgggcaa gctggctgag tacggactga tccctctgtt catcccctac 1200

accgacagca acgagcccat cgtgaaagaa atcaagtgga tggaaaagtc ccggaaccag 1260

agcgtgcggc ggctggataa ggacatgttc attcaggccc tggaacggtt cctgagctgg 1320

gagagctgga acctgaaagt gaaagaggaa tacgagaagg tcgagaaaga gtacaagacc 1380

ctggaagaga ggatcaaaga ggacatccag gctctgaagg ctctggaaca gtatgagaaa 1440

gagcggcaag aacagctgct gcgggacacc ctgaacacca acgagtaccg gctgagcaag 1500

agaggcctta gaggctggcg ggaaatcatc cagaaatggc tgaaaatgga cgagaacgag 1560

ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc tagagaggcc 1620

ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat ctggcggaat 1680

caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa aaagaaggac 1740

gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct gtgggtccga 1800

ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga gcagctgcac 1860

accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta ccctacagaa 1920

tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag ccggcagttc 1980

tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac ctacaaggat 2040

gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca gttcgacaga 2100

gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag aatctacttc 2160

aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct gaagatccac 2220

cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga gtggatcaag 2280

gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg cctgagagtg 2340

atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga ggtggtggat 2400

cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga gctgtatgcc 2460

gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa gagcagagaa 2520

gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct caacttcctg 2580

cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa gcgggtcacc 2640

aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga tgagctgatc 2700

cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct gaagcagctc 2760

cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa gtccctgagc 2820

gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat cgatcggacc 2880

cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt gcgtagactg 2940

gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct gaaagaagat 3000

cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg ctacgacgtg 3060

cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt cgaggatctg 3120

agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct catgaagtgg 3180

tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg cctgcaagtg 3240

ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag ccctggcatc 3300

agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa gaatctgcag 3360

agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga tctgtaccca 3420

gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt gaccacacac 3480

gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac ccacggcttc 3540

tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 3
<211> 1361
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at P153 polypeptide sequence

<400> 3

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp

```
145                 150                 155                 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Gly Gly Ser Gly Gly Ser Ser
                165                 170                 175

Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu
            180                 185                 190

Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu
        195                 200                 205

Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly
    210                 215                 220

Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly
225                 230                 235                 240

Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val
                245                 250                 255

Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg
            260                 265                 270

Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala
        275                 280                 285

Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg Val
    290                 295                 300

Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys
305                 310                 315                 320

Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala
                325                 330                 335

Gln Ser Ser Thr Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
            340                 345                 350

Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Trp Glu Glu Glu Lys Lys
        355                 360                 365
```

```
Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
    370                 375                 380

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
385                 390                 395                 400

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
                405                 410                 415

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
            420                 425                 430

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
        435                 440                 445

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
    450                 455                 460

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
465                 470                 475                 480

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
                485                 490                 495

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
            500                 505                 510

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
        515                 520                 525

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
    530                 535                 540

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
545                 550                 555                 560

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
                565                 570                 575

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
```

```
            580                 585                 590

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
        595                 600                 605

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
    610                 615                 620

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
625                 630                 635                 640

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
                645                 650                 655

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
            660                 665                 670

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
        675                 680                 685

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
    690                 695                 700

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
705                 710                 715                 720

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
                725                 730                 735

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
            740                 745                 750

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
        755                 760                 765

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
    770                 775                 780

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
785                 790                 795                 800
```

```
Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
                805                 810                 815

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
            820                 825                 830

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
        835                 840                 845

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
    850                 855                 860

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
865                 870                 875                 880

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
                885                 890                 895

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
            900                 905                 910

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
        915                 920                 925

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
    930                 935                 940

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
945                 950                 955                 960

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
                965                 970                 975

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
            980                 985                 990

Asn Ala Leu Lys Glu Asp Arg Leu  Lys Lys Met Ala Asn  Thr Ile Ile
        995                 1000                1005

Met His  Ala Leu Gly Tyr Cys  Tyr Asp Val Arg Lys  Lys Lys Trp
```

```
    1010                1015                1020

Gln Ala  Lys Asn Pro Ala Cys  Gln Ile Ile Leu Phe  Glu Asp Leu
    1025                1030                1035

Ser Asn  Tyr Asn Pro Tyr Glu  Glu Arg Ser Arg Phe  Glu Asn Ser
    1040                1045                1050

Lys Leu  Met Lys Trp Ser Arg  Arg Glu Ile Pro Arg  Gln Val Ala
    1055                1060                1065

Leu Gln  Gly Glu Ile Tyr Gly  Leu Gln Val Gly Glu  Val Gly Ala
    1070                1075                1080

Gln Phe  Ser Ser Arg Phe His  Ala Lys Thr Gly Ser  Pro Gly Ile
    1085                1090                1095

Arg Cys  Ser Val Val Thr Lys  Glu Lys Leu Gln Asp  Asn Arg Phe
    1100                1105                1110

Phe Lys  Asn Leu Gln Arg Glu  Gly Arg Leu Thr Leu  Asp Lys Ile
    1115                1120                1125

Ala Val  Leu Lys Glu Gly Asp  Leu Tyr Pro Asp Lys  Gly Gly Glu
    1130                1135                1140

Lys Phe  Ile Ser Leu Ser Lys  Asp Arg Lys Cys Val  Thr Thr His
    1145                1150                1155

Ala Asp  Ile Asn Ala Ala Gln  Asn Leu Gln Lys Arg  Phe Trp Thr
    1160                1165                1170

Arg Thr  His Gly Phe Tyr Lys  Val Tyr Cys Lys Ala  Tyr Gln Val
    1175                1180                1185

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

Lys Ile  Ile Glu Glu Phe Gly  Glu Gly Tyr Phe Ile  Leu Lys Asp
    1205                1210                1215
```

```
Gly Val  Tyr Glu Trp Val Asn  Ala Gly Lys Leu Lys  Ile Lys Lys
    1220                1225                1230


Gly Ser  Ser Lys Gln Ser Ser  Ser Glu Leu Val Asp  Ser Asp Ile
    1235                1240                1245


Leu Lys  Asp Ser Phe Asp Leu  Ala Ser Glu Leu Lys  Gly Glu Lys
    1250                1255                1260


Leu Met  Leu Tyr Arg Asp Pro  Ser Gly Asn Val Phe  Pro Ser Asp
    1265                1270                1275


Lys Trp  Met Ala Ala Gly Val  Phe Phe Gly Lys Leu  Glu Arg Ile
    1280                1285                1290


Leu Ile  Ser Lys Leu Thr Asn  Gln Tyr Ser Ile Ser  Thr Ile Glu
    1295                1300                1305


Asp Asp  Ser Ser Lys Gln Ser  Met Lys Arg Pro Ala  Ala Thr Lys
    1310                1315                1320


Lys Ala  Gly Gln Ala Lys Lys  Lys Lys Gly Ser Tyr  Pro Tyr Asp
    1325                1330                1335


Val Pro  Asp Tyr Ala Tyr Pro  Tyr Asp Val Pro Asp  Tyr Ala Tyr
    1340                1345                1350


Pro Tyr  Asp Val Pro Asp Tyr  Ala
    1355                1360


<210>  4
<211>  4092
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20
       at K255 polynucleotide sequence


<220>
<221>  misc_feature
<222>  (1)..(6)
```

<223> Kozak sequence

<220>
<221> misc_feature
<222> (7)..(57)
<223> Sequence encoding N-terminal nuclear localization signal (NLS)

<220>
<221> misc_feature
<222> (58)..(822)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (823)..(840)
<223> Sequence encoding a linker

<220>
<221> misc_feature
<222> (841)..(1338)
<223> Sequence encoding TadA*8

<220>
<221> misc_feature
<222> (1339)..(1398)
<223> Sequence encoding Xten20 linker

<220>
<221> misc_feature
<222> (1399)..(3954)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (3955)..(4002)
<223> Sequence encoding C-terminal NLS

<220>
<221> misc_feature
<222> (4003)..(4008)
<223> Sequence encoding a GS linker

<220>
<221> misc_feature
<222> (4009)..(4089)
<223> Sequence encoding 3x hemagglutinin (HA) tag

<400> 4
gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aaggaggctc tggaggaagc 840

tccgaagtcg agttttccca tgagtactgg atgagacacg cattgactct cgcaaagagg 900

gctcgagatg aacgcgaggt gcccgtgggg gcagtactcg tgctcaacaa tcgcgtaatc 960

ggcgaaggtt ggaatagggc aatcggactc cacgacccca ctgcacatgc ggaaatcatg 1020

gcccttcgac agggagggct tgtgatgcag aattatcgac tttatgatgc gacgctgtac 1080

gtcacgtttg aaccttgcgt aatgtgcgcg ggagctatga ttcactcccg cattggacga 1140

gttgtattcg gtgttcgcaa cgccaagacg ggtgccgcag gttcactgat ggacgtgctg 1200

catcatccag gcatgaacca ccgggtagaa atcacagaag gcatattggc ggacgaatgt 1260

gcggcgctgt tgtgtcgttt ttttcgcatg cccaggcggg tctttaacgc ccagaaaaaa 1320

gcacaatcct ctactgacgg ctcttctgga tctgaaacac ctggcacaag cgagagcgcc 1380

acccctgaga gctctggcga ggacatccag gctctgaagg ctctggaaca gtatgagaaa 1440

gagcggcaag aacagctgct gcgggacacc ctgaacacca acgagtaccg gctgagcaag 1500

agaggcctta gaggctggcg ggaaatcatc cagaaatggc tgaaaatgga cgagaacgag 1560

ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc tagagaggcc 1620

ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat ctggcggaat 1680

caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa aaagaaggac 1740

gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct gtgggtccga 1800

ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga gcagctgcac 1860

accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta ccctacagaa 1920

tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag ccggcagttc 1980

tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac ctacaaggat 2040

gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca gttcgacaga 2100

gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag aatctacttc 2160

aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct gaagatccac 2220

cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga gtggatcaag 2280

gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg cctgagagtg 2340

atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga ggtggtggat 2400

cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga gctgtatgcc 2460

gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa gagcagagaa 2520

gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct caacttcctg 2580

cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa gcgggtcacc 2640

aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga tgagctgatc 2700

cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct gaagcagctc 2760

cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa gtccctgagc 2820

gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat cgatcggacc 2880

cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt gcgtagactg 2940

gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct gaaagaagat 3000

cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg ctacgacgtg 3060

cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt cgaggatctg 3120

agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct catgaagtgg 3180

tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg cctgcaagtg 3240

ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag ccctggcatc 3300

agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa gaatctgcag 3360

agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga tctgtaccca 3420

gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt gaccacacac 3480

gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac ccacggcttc 3540

tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 5
<211> 1361
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at K255 polypeptide sequence

<400> 5

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

```
Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
                85                  90                  95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
            100                 105                 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
        115                 120                 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
    130                 135                 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145                 150                 155                 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
                165                 170                 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
            180                 185                 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
        195                 200                 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
    210                 215                 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
225                 230                 235                 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
                245                 250                 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
            260                 265                 270

Gly Gly Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr Trp
        275                 280                 285
```

```
Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu
    290                 295                 300

Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu
305                 310                 315                 320

Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu
                325                 330                 335

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu
            340                 345                 350

Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala
        355                 360                 365

Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg
    370                 375                 380

Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His His
385                 390                 395                 400

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp
                405                 410                 415

Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val
            420                 425                 430

Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Ser Ser Gly
        435                 440                 445

Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly
    450                 455                 460

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
465                 470                 475                 480

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
                485                 490                 495

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
            500                 505                 510
```

```
Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
        515                 520                 525

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
    530                 535                 540

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
545                 550                 555                 560

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
                565                 570                 575

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
            580                 585                 590

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
        595                 600                 605

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
    610                 615                 620

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
625                 630                 635                 640

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
                645                 650                 655

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
            660                 665                 670

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
        675                 680                 685

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
    690                 695                 700

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
705                 710                 715                 720
```

```
Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
                725                 730                 735

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
            740                 745                 750

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
        755                 760                 765

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
    770                 775                 780

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
785                 790                 795                 800

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
                805                 810                 815

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
            820                 825                 830

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
        835                 840                 845

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
    850                 855                 860

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
865                 870                 875                 880

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
                885                 890                 895

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
            900                 905                 910

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
        915                 920                 925

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
    930                 935                 940
```

```
Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
945                 950                 955                 960


Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
                965                 970                 975


Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
            980                 985                 990


Asn Ala Leu Lys Glu Asp Arg Leu  Lys Lys Met Ala Asn  Thr Ile Ile
        995                 1000                1005


Met His  Ala Leu Gly Tyr Cys  Tyr Asp Val Arg Lys  Lys Lys Trp
    1010                1015                1020


Gln Ala  Lys Asn Pro Ala Cys  Gln Ile Ile Leu Phe  Glu Asp Leu
    1025                1030                1035


Ser Asn  Tyr Asn Pro Tyr Glu  Glu Arg Ser Arg Phe  Glu Asn Ser
    1040                1045                1050


Lys Leu  Met Lys Trp Ser Arg  Arg Glu Ile Pro Arg  Gln Val Ala
    1055                1060                1065


Leu Gln  Gly Glu Ile Tyr Gly  Leu Gln Val Gly Glu  Val Gly Ala
    1070                1075                1080


Gln Phe  Ser Ser Arg Phe His  Ala Lys Thr Gly Ser  Pro Gly Ile
    1085                1090                1095


Arg Cys  Ser Val Val Thr Lys  Glu Lys Leu Gln Asp  Asn Arg Phe
    1100                1105                1110


Phe Lys  Asn Leu Gln Arg Glu  Gly Arg Leu Thr Leu  Asp Lys Ile
    1115                1120                1125


Ala Val  Leu Lys Glu Gly Asp  Leu Tyr Pro Asp Lys  Gly Gly Glu
    1130                1135                1140
```

```
Lys Phe  Ile Ser Leu Ser Lys  Asp Arg Lys Cys Val  Thr Thr His
    1145                1150                1155

Ala Asp  Ile Asn Ala Ala Gln  Asn Leu Gln Lys Arg  Phe Trp Thr
    1160                1165                1170

Arg Thr  His Gly Phe Tyr Lys  Val Tyr Cys Lys Ala  Tyr Gln Val
    1175                1180                1185

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

Lys Ile  Ile Glu Glu Phe Gly  Glu Gly Tyr Phe Ile  Leu Lys Asp
    1205                1210                1215

Gly Val  Tyr Glu Trp Val Asn  Ala Gly Lys Leu Lys  Ile Lys Lys
    1220                1225                1230

Gly Ser  Ser Lys Gln Ser Ser  Ser Glu Leu Val Asp  Ser Asp Ile
    1235                1240                1245

Leu Lys  Asp Ser Phe Asp Leu  Ala Ser Glu Leu Lys  Gly Glu Lys
    1250                1255                1260

Leu Met  Leu Tyr Arg Asp Pro  Ser Gly Asn Val Phe  Pro Ser Asp
    1265                1270                1275

Lys Trp  Met Ala Ala Gly Val  Phe Phe Gly Lys Leu  Glu Arg Ile
    1280                1285                1290

Leu Ile  Ser Lys Leu Thr Asn  Gln Tyr Ser Ile Ser  Thr Ile Glu
    1295                1300                1305

Asp Asp  Ser Ser Lys Gln Ser  Met Lys Arg Pro Ala  Ala Thr Lys
    1310                1315                1320

Lys Ala  Gly Gln Ala Lys Lys  Lys Lys Gly Ser Tyr  Pro Tyr Asp
    1325                1330                1335

Val Pro  Asp Tyr Ala Tyr Pro  Tyr Asp Val Pro Asp  Tyr Ala Tyr
    1340                1345                1350
```

Pro Tyr Asp Val Pro Asp Tyr Ala
    1355   1360

<210> 6
<211> 4092
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20
       at D306 polynucleotide sequence

<220>
<221> misc_feature
<222> (1)..(6)
<223> Kozak sequence

<220>
<221> misc_feature
<222> (7)..(57)
<223> Sequence encoding N-terminal nuclear localization signal (NLS)

<220>
<221> misc_feature
<222> (58)..(975)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (976)..(993)
<223> Sequence encoding a linker

<220>
<221> misc_feature
<222> (994)..(1491)
<223> Sequence encoding TadA*8

<220>
<221> misc_feature
<222> (1492)..(1551)
<223> Sequence encoding Xten20 linker

<220>
<221> misc_feature
<222> (1552)..(3954)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (3955)..(4002)

<223> Sequence encoding C-terminal NLS

<220>
<221> misc_feature
<222> (4003)..(4008)
<223> Sequence encoding a GS linker

<220>
<221> misc_feature
<222> (4009)..(4089)
<223> Sequence encoding 3x hemagglutinin (HA) tag

<400> 6
gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aagaggacat ccaggctctg 840

aaggctctgg aacagtatga gaaagagcgg caagaacagc tgctgcggga caccctgaac 900

accaacgagt accggctgag caagagaggc cttagaggct ggcgggaaat catccagaaa 960

tggctgaaaa tggacggagg ctctggagga agctccgaag tcgagttttc ccatgagtac 1020

tggatgagac acgcattgac tctcgcaaag agggctcgag atgaacgcga ggtgcccgtg 1080

ggggcagtac tcgtgctcaa caatcgcgta atcggcgaag gttggaatag ggcaatcgga 1140

ctccacgacc ccactgcaca tgcggaaatc atggcccttc gacagggagg gcttgtgatg 1200

cagaattatc gactttatga tgcgacgctg tacgtcacgt ttgaaccttg cgtaatgtgc 1260

gcgggagcta tgattcactc ccgcattgga cgagttgtat tcggtgttcg caacgccaag 1320

acgggtgccg caggttcact gatggacgtg ctgcatcatc caggcatgaa ccaccgggta 1380

gaaatcacag aaggcatatt ggcggacgaa tgtgcggcgc tgttgtgtcg ttttttttcgc 1440

atgcccaggc gggtctttaa cgcccagaaa aaagcacaat cctctactga cggctcttct 1500

ggatctgaaa cacctggcac aagcgagagc gccacccctg agagctctgg cgagaacgag 1560

ccctccgaga agtacctgga agtgttcaag gactaccagc ggaagcaccc tagagaggcc 1620

ggcgattaca gcgtgtacga gttcctgtcc aagaaagaga accacttcat ctggcggaat 1680

caccctgagt acccctacct gtacgccacc ttctgcgaga tcgacaagaa aaagaaggac 1740

gccaagcagc aggccacctt cacactggcc gatcctatca atcaccctct gtgggtccga 1800

ttcgaggaaa gaagcggcag caacctgaac aagtacagaa tcctgaccga gcagctgcac 1860

accgagaagc tgaagaaaaa gctgacagtg cagctggacc ggctgatcta ccctacagaa 1920

tctggcggct gggaagagaa gggcaaagtg gacattgtgc tgctgcccag ccggcagttc 1980

tacaaccaga tcttcctgga catcgaggaa aagggcaagc acgccttcac ctacaaggat 2040

gagagcatca agttccctct gaagggcaca ctcggcggag ccagagtgca gttcgacaga 2100

gatcacctga gaagataccc tcacaaggtg gaaagcggca acgtgggcag aatctacttc 2160

aacatgaccg tgaacatcga gcctacagag tccccagtgt ccaagtctct gaagatccac 2220

cgggacgact tccccaaggt ggtcaacttc aagcccaaag aactgaccga gtggatcaag 2280

gacagcaagg gcaagaaact gaagtccggc atcgagtccc tggaaatcgg cctgagagtg 2340

atgagcatcg acctgggaca gagacaggcc gctgccgcct ctattttcga ggtggtggat 2400

cagaagcccg acatcgaagg caagctgttt ttcccaatca agggcaccga gctgtatgcc 2460

gtgcacagag ccagcttcaa catcaagctg cccggcgaga cactggtcaa gagcagagaa 2520

gtgctgcgga aggccagaga ggacaatctg aaactgatga accagaagct caacttcctg 2580

cggaacgtgc tgcacttcca gcagttcgag gacatcaccg agagagagaa gcgggtcacc 2640

aagtggatca gcagacaaga gaacagcgac gtgcccctgg tgtaccagga tgagctgatc 2700

cagatccgcg agctgatgta caagccttac aaggactggg tcgccttcct gaagcagctc 2760

cacaagagac tggaagtcga gatcggcaaa gaagtgaagc actggcggaa gtccctgagc 2820

gacggaagaa agggcctgta cggcatctcc ctgaagaaca tcgacgagat cgatcggacc 2880

cggaagttcc tgctgagatg gtccctgagg cctaccgaac ctggcgaagt gcgtagactg 2940

gaacccggcc agagattcgc catcgaccag ctgaatcacc tgaacgccct gaaagaagat 3000

cggctgaaga agatggccaa caccatcatc atgcacgccc tgggctactg ctacgacgtg 3060

cggaagaaga aatggcaggc taagaacccc gcctgccaga tcatcctgtt cgaggatctg 3120

agcaactaca acccctacga ggaaaggtcc cgcttcgaga acagcaagct catgaagtgg 3180

tccagacgcg agatccccag acaggttgca ctgcagggcg agatctatgg cctgcaagtg 3240

ggagaagtgg gcgctcagtt cagcagcaga ttccacgcca agacaggcag ccctggcatc 3300

agatgtagcg tcgtgaccaa agagaagctg caggacaatc ggttcttcaa gaatctgcag 3360

agagagggca gactgaccct ggacaaaatc gccgtgctga aagagggcga tctgtaccca 3420

gacaaaggcg gcgagaagtt catcagcctg agcaaggatc ggaagtgcgt gaccacacac 3480

gccgacatca acgccgctca gaacctgcag aagcggttct ggacaagaac ccacggcttc 3540

tacaaggtgt actgcaaggc ctaccaggtg gacggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 7
<211> 1361
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at D306 polypeptide sequence

<400> 7

```
Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1               5                  10                  15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
            20                  25                  30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
        35                  40                  45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
    50                  55                  60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65                  70                  75                  80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
                85                  90                  95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
            100                 105                 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
        115                 120                 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
    130                 135                 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145                 150                 155                 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
                165                 170                 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
            180                 185                 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
        195                 200                 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
```

```
    210                 215                 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
225                 230                 235                 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
                245                 250                 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
            260                 265                 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
        275                 280                 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
    290                 295                 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
305                 310                 315                 320

Lys Met Asp Gly Gly Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His
                325                 330                 335

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp
            340                 345                 350

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val
        355                 360                 365

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala
    370                 375                 380

His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn
385                 390                 395                 400

Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val
                405                 410                 415

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe
            420                 425                 430
```

```
Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val
        435                 440                 445

Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile
    450                 455                 460

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro
465                 470                 475                 480

Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly
                485                 490                 495

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
            500                 505                 510

Ser Ser Gly Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
        515                 520                 525

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
    530                 535                 540

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
545                 550                 555                 560

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
                565                 570                 575

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
            580                 585                 590

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
        595                 600                 605

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
    610                 615                 620

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
625                 630                 635                 640

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
```

```
               645                 650                 655

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
            660                 665                 670

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
        675                 680                 685

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
    690                 695                 700

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
705                 710                 715                 720

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
                725                 730                 735

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
            740                 745                 750

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
        755                 760                 765

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
    770                 775                 780

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
785                 790                 795                 800

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
                805                 810                 815

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
            820                 825                 830

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
        835                 840                 845

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
    850                 855                 860
```

```
Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
865                 870                 875                 880


Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
                885                 890                 895


Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
            900                 905                 910


Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
        915                 920                 925


Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
    930                 935                 940


Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
945                 950                 955                 960


Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
                965                 970                 975


Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
            980                 985                 990


Asn Ala Leu Lys Glu Asp Arg Leu  Lys Lys Met Ala Asn  Thr Ile Ile
        995                 1000                1005


Met His  Ala Leu Gly Tyr Cys  Tyr Asp Val Arg Lys  Lys Lys Trp
    1010                1015                1020


Gln Ala  Lys Asn Pro Ala Cys  Gln Ile Ile Leu Phe  Glu Asp Leu
    1025                1030                1035


Ser Asn  Tyr Asn Pro Tyr Glu  Glu Arg Ser Arg Phe  Glu Asn Ser
    1040                1045                1050


Lys Leu  Met Lys Trp Ser Arg  Arg Glu Ile Pro Arg  Gln Val Ala
    1055                1060                1065


Leu Gln  Gly Glu Ile Tyr Gly  Leu Gln Val Gly Glu  Val Gly Ala
```

```
    1070                1075                1080

Gln Phe  Ser Ser Arg Phe His  Ala Lys Thr Gly Ser  Pro Gly Ile
    1085                1090                1095

Arg Cys  Ser Val Val Thr Lys  Glu Lys Leu Gln Asp  Asn Arg Phe
    1100                1105                1110

Phe Lys  Asn Leu Gln Arg Glu  Gly Arg Leu Thr Leu  Asp Lys Ile
    1115                1120                1125

Ala Val  Leu Lys Glu Gly Asp  Leu Tyr Pro Asp Lys  Gly Gly Glu
    1130                1135                1140

Lys Phe  Ile Ser Leu Ser Lys  Asp Arg Lys Cys Val  Thr Thr His
    1145                1150                1155

Ala Asp  Ile Asn Ala Ala Gln  Asn Leu Gln Lys Arg  Phe Trp Thr
    1160                1165                1170

Arg Thr  His Gly Phe Tyr Lys  Val Tyr Cys Lys Ala  Tyr Gln Val
    1175                1180                1185

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

Lys Ile  Ile Glu Glu Phe Gly  Glu Gly Tyr Phe Ile  Leu Lys Asp
    1205                1210                1215

Gly Val  Tyr Glu Trp Val Asn  Ala Gly Lys Leu Lys  Ile Lys Lys
    1220                1225                1230

Gly Ser  Ser Lys Gln Ser Ser  Ser Glu Leu Val Asp  Ser Asp Ile
    1235                1240                1245

Leu Lys  Asp Ser Phe Asp Leu  Ala Ser Glu Leu Lys  Gly Glu Lys
    1250                1255                1260

Leu Met  Leu Tyr Arg Asp Pro  Ser Gly Asn Val Phe  Pro Ser Asp
    1265                1270                1275
```

Lys Trp Met Ala Ala Gly Val Phe Phe Gly Lys Leu Glu Arg Ile
1280 1285 1290

Leu Ile Ser Lys Leu Thr Asn Gln Tyr Ser Ile Ser Thr Ile Glu
1295 1300 1305

Asp Asp Ser Ser Lys Gln Ser Met Lys Arg Pro Ala Ala Thr Lys
1310 1315 1320

Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp
1325 1330 1335

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
1340 1345 1350

Pro Tyr Asp Val Pro Asp Tyr Ala
1355 1360

<210> 8
<211> 4092
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at D980 polynucleotide sequence

<220>
<221> misc_feature
<222> (1)..(6)
<223> Kozak sequence

<220>
<221> misc_feature
<222> (7)..(57)
<223> Sequence encoding N-terminal nuclear localization signal (NLS)

<220>
<221> misc_feature
<222> (58)..(2997)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (2998)..(3015)
<223> Sequence encoding a linker

<220>
<221> misc_feature
<222> (3016)..(3513)
<223> Sequence encoding TadA*8

<220>
<221> misc_feature
<222> (3514)..(3573)
<223> Sequence encoding Xten20 linker

<220>
<221> misc_feature
<222> (3574)..(3954)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (3955)..(4002)
<223> Sequence encoding C-terminal NLS

<220>
<221> misc_feature
<222> (4003)..(4008)
<223> Sequence encoding a GS linker

<220>
<221> misc_feature
<222> (4009)..(4089)
<223> Sequence encoding 3x hemagglutinin (HA) tag

<400> 8
gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aagaggacat ccaggctctg 840

aaggctctgg aacagtatga gaaagagcgg caagaacagc tgctgcggga caccctgaac 900

accaacgagt accggctgag caagagaggc cttagaggct ggcgggaaat catccagaaa 960

tggctgaaaa tggacgagaa cgagccctcc gagaagtacc tggaagtgtt caaggactac 1020

cagcggaagc accctagaga ggccggcgat tacagcgtgt acgagttcct gtccaagaaa 1080

gagaaccact tcatctggcg gaatcaccct gagtacccct acctgtacgc caccttctgc 1140

gagatcgaca agaaaaagaa ggacgccaag cagcaggcca ccttcacact ggccgatcct 1200

atcaatcacc ctctgtgggt ccgattcgag gaaagaagcg gcagcaacct gaacaagtac 1260

agaatcctga ccgagcagct gcacaccgag aagctgaaga aaaagctgac agtgcagctg 1320

gaccggctga tctaccctac agaatctggc ggctgggaag agaagggcaa agtggacatt 1380

gtgctgctgc ccagccggca gttctacaac cagatcttcc tggacatcga ggaaaagggc 1440

aagcacgcct tcacctacaa ggatgagagc atcaagttcc ctctgaaggg cacactcggc 1500

ggagccagag tgcagttcga cagagatcac ctgagaagat accctcacaa ggtggaaagc 1560

ggcaacgtgg gcagaatcta cttcaacatg accgtgaaca tcgagcctac agagtcccca 1620

gtgtccaagt ctctgaagat ccaccgggac gacttcccca aggtggtcaa cttcaagccc 1680

aaagaactga ccgagtggat caaggacagc aagggcaaga aactgaagtc cggcatcgag 1740

tccctggaaa tcggcctgag agtgatgagc atcgacctgg gacagagaca ggccgctgcc 1800

gcctctattt tcgaggtggt ggatcagaag cccgacatcg aaggcaagct gtttttccca 1860

atcaagggca ccgagctgta tgccgtgcac agagccagct tcaacatcaa gctgcccggc 1920

gagacactgg tcaagagcag agaagtgctg cggaaggcca gagaggacaa tctgaaactg 1980

atgaaccaga agctcaactt cctgcggaac gtgctgcact tccagcagtt cgaggacatc 2040

accgagagag agaagcgggt caccaagtgg atcagcagac aagagaacag cgacgtgccc 2100

ctggtgtacc aggatgagct gatccagatc cgcgagctga tgtacaagcc ttacaaggac 2160

tgggtcgcct tcctgaagca gctccacaag agactggaag tcgagatcgg caaagaagtg 2220

aagcactggc ggaagtccct gagcgacgga agaaagggcc tgtacggcat ctccctgaag 2280

aacatcgacg agatcgatcg gacccggaag ttcctgctga gatggtccct gaggcctacc 2340

gaacctggcg aagtgcgtag actggaaccc ggccagagat tcgccatcga ccagctgaat 2400

cacctgaacg ccctgaaaga agatcggctg aagaagatgg ccaacaccat catcatgcac 2460

gccctgggct actgctacga cgtgcggaag aagaaatggc aggctaagaa ccccgcctgc 2520

cagatcatcc tgttcgagga tctgagcaac tacaacccct acgaggaaag gtcccgcttc 2580

gagaacagca agctcatgaa gtggtccaga cgcgagatcc ccagacaggt tgcactgcag 2640

ggcgagatct atggcctgca agtgggagaa gtgggcgctc agttcagcag cagattccac 2700

gccaagacag gcagccctgg catcagatgt agcgtcgtga ccaaagagaa gctgcaggac 2760

aatcggttct tcaagaatct gcagagagag ggcagactga ccctggacaa aatcgccgtg 2820

ctgaaagagg gcgatctgta cccagacaaa ggcggcgaga agttcatcag cctgagcaag 2880

gatcggaagt gcgtgaccac acacgccgac atcaacgccg ctcagaacct gcagaagcgg 2940

ttctggacaa gaacccacgg cttctacaag gtgtactgca aggcctacca ggtggacgga 3000

ggctctggag gaagctccga agtcgagttt tcccatgagt actggatgag acacgcattg 3060

actctcgcaa agagggctcg agatgaacgc gaggtgcccg tgggggcagt actcgtgctc 3120

aacaatcgcg taatcggcga aggttggaat agggcaatcg gactccacga ccccactgca 3180

catgcggaaa tcatggccct tcgacaggga gggcttgtga tgcagaatta tcgactttat 3240

gatgcgacgc tgtacgtcac gtttgaacct tgcgtaatgt gcgcgggagc tatgattcac 3300

tcccgcattg gacgagttgt attcggtgtt cgcaacgcca agacgggtgc cgcaggttca 3360

ctgatggacg tgctgcatca tccaggcatg aaccaccggg tagaaatcac agaaggcata 3420

ttggcggacg aatgtgcggc gctgttgtgt cgtttttttc gcatgcccag gcgggtcttt 3480

aacgcccaga aaaaagcaca atcctctact gacggctctt ctggatctga aacacctggc 3540

acaagcgaga gcgccacccc tgagagctct ggcggccaga ccgtgtacat ccctgagagc 3600

aaggaccaga agcagaagat catcgaagag ttcggcgagg gctacttcat tctgaaggac 3660

ggggtgtacg aatgggtcaa cgccggcaag ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 9
<211> 1361
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at D980 polypeptide sequence

<400> 9

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65 70 75 80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
85 90 95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
100 105 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
115 120 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
130 135 140

```
Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145                 150                 155                 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
                165                 170                 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
            180                 185                 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
        195                 200                 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
    210                 215                 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
225                 230                 235                 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
                245                 250                 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
            260                 265                 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
        275                 280                 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
    290                 295                 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
305                 310                 315                 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
                325                 330                 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
            340                 345                 350
```

```
Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
        355                 360                 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
    370                 375                 380

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
385                 390                 395                 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
                405                 410                 415

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
            420                 425                 430

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
        435                 440                 445

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
    450                 455                 460

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

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
                485                 490                 495

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
            500                 505                 510

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
        515                 520                 525

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
    530                 535                 540

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
545                 550                 555                 560

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
                565                 570                 575
```

```
Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
            580                 585                 590

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
        595                 600                 605

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
    610                 615                 620

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
625                 630                 635                 640

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
                645                 650                 655

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
            660                 665                 670

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
        675                 680                 685

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
    690                 695                 700

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
705                 710                 715                 720

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
                725                 730                 735

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
            740                 745                 750

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
        755                 760                 765

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
    770                 775                 780
```

```
Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
785                 790                 795                 800


Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
                805                 810                 815


Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
            820                 825                 830


Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
        835                 840                 845


Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
    850                 855                 860


Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
865                 870                 875                 880


Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
                885                 890                 895


Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
            900                 905                 910


Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
        915                 920                 925


Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
    930                 935                 940


Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
945                 950                 955                 960


Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
                965                 970                 975


Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
            980                 985                 990


Ala Tyr Gln Val Asp Gly Gly Ser  Gly Gly Ser Ser Glu  Val Glu Phe
        995                 1000                1005
```

```
Ser His  Glu Tyr Trp Met Arg  His Ala Leu Thr Leu  Ala Lys Arg
    1010                1015                1020

Ala Arg  Asp Glu Arg Glu Val  Pro Val Gly Ala Val  Leu Val Leu
    1025                1030                1035

Asn Asn  Arg Val Ile Gly Glu  Gly Trp Asn Arg Ala  Ile Gly Leu
    1040                1045                1050

His Asp  Pro Thr Ala His Ala  Glu Ile Met Ala Leu  Arg Gln Gly
    1055                1060                1065

Gly Leu  Val Met Gln Asn Tyr  Arg Leu Tyr Asp Ala  Thr Leu Tyr
    1070                1075                1080

Val Thr  Phe Glu Pro Cys Val  Met Cys Ala Gly Ala  Met Ile His
    1085                1090                1095

Ser Arg  Ile Gly Arg Val Val  Phe Gly Val Arg Asn  Ala Lys Thr
    1100                1105                1110

Gly Ala  Ala Gly Ser Leu Met  Asp Val Leu His His  Pro Gly Met
    1115                1120                1125

Asn His  Arg Val Glu Ile Thr  Glu Gly Ile Leu Ala  Asp Glu Cys
    1130                1135                1140

Ala Ala  Leu Leu Cys Arg Phe  Phe Arg Met Pro Arg  Arg Val Phe
    1145                1150                1155

Asn Ala  Gln Lys Lys Ala Gln  Ser Ser Thr Asp Gly  Ser Ser Gly
    1160                1165                1170

Ser Glu  Thr Pro Gly Thr Ser  Glu Ser Ala Thr Pro  Glu Ser Ser
    1175                1180                1185

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

```
Lys Ile  Ile Glu Glu Phe Gly  Glu Gly Tyr Phe Ile  Leu Lys Asp
    1205                 1210                 1215

Gly Val  Tyr Glu Trp Val Asn  Ala Gly Lys Leu Lys  Ile Lys Lys
    1220                 1225                 1230

Gly Ser  Ser Lys Gln Ser Ser  Ser Glu Leu Val Asp  Ser Asp Ile
    1235                 1240                 1245

Leu Lys  Asp Ser Phe Asp Leu  Ala Ser Glu Leu Lys  Gly Glu Lys
    1250                 1255                 1260

Leu Met  Leu Tyr Arg Asp Pro  Ser Gly Asn Val Phe  Pro Ser Asp
    1265                 1270                 1275

Lys Trp  Met Ala Ala Gly Val  Phe Phe Gly Lys Leu  Glu Arg Ile
    1280                 1285                 1290

Leu Ile  Ser Lys Leu Thr Asn  Gln Tyr Ser Ile Ser  Thr Ile Glu
    1295                 1300                 1305

Asp Asp  Ser Ser Lys Gln Ser  Met Lys Arg Pro Ala  Ala Thr Lys
    1310                 1315                 1320

Lys Ala  Gly Gln Ala Lys Lys  Lys Lys Gly Ser Tyr  Pro Tyr Asp
    1325                 1330                 1335

Val Pro  Asp Tyr Ala Tyr Pro  Tyr Asp Val Pro Asp  Tyr Ala Tyr
    1340                 1345                 1350

Pro Tyr  Asp Val Pro Asp Tyr  Ala
    1355                 1360
```

<210> 10
<211> 4092
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at K1019 polynucleotide sequence

<220>
<221> misc_feature
<222> (1)..(6)
<223> Kozak sequence

<220>
<221> misc_feature
<222> (7)..(57)
<223> Sequence encoding N-terminal nuclear localization signal (NLS)

<220>
<221> misc_feature
<222> (58)..(3114)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (3115)..(3132)
<223> Sequence encoding a linker

<220>
<221> misc_feature
<222> (3133)..(3630)
<223> Sequence encoding TadA*8

<220>
<221> misc_feature
<222> (3631)..(3690)
<223> Sequence encoding Xten20 linker

<220>
<221> misc_feature
<222> (3691)..(3954)
<223> Sequence encoding BhCas12b

<220>
<221> misc_feature
<222> (3955)..(4002)
<223> Sequence encoding C-terminal NLS

<220>
<221> misc_feature
<222> (4003)..(4008)
<223> Sequence encoding GS linker

<220>
<221> misc_feature
<222> (4009)..(4089)
<223> Sequence encoding 3x hemagglutinin (HA) tag

<400> 10
gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccgcc 60

accagatcct tcatcctgaa gatcgagccc aacgaggaag tgaagaaagg cctctggaaa 120

acccacgagg tgctgaacca cggaatcgcc tactacatga atatcctgaa gctgatccgg 180

caagaggcca tctacgagca ccacgagcag gaccccaaga atcccaagaa ggtgtccaag 240

gccgagatcc aggccgagct gtgggatttc gtgctgaaga tgcagaagtg caacagcttc 300

acacacgagg tggacaagga cgaggtgttc aacatcctga gagagctgta cgaggaactg 360

gtgcccagca gcgtggaaaa gaagggcgaa gccaaccagc tgagcaacaa gtttctgtac 420

cctctggtgg accccaacag ccagtctgga aagggaacag ccagcagcgg cagaaagccc 480

agatggtaca acctgaagat tgccggcgat ccctcctggg aagaagagaa gaagaagtgg 540

gaagaagata agaaaaagga cccgctggcc aagatcctgg gcaagctggc tgagtacgga 600

ctgatccctc tgttcatccc ctacaccgac agcaacgagc ccatcgtgaa agaaatcaag 660

tggatggaaa agtcccggaa ccagagcgtg cggcggctgg ataaggacat gttcattcag 720

gccctggaac ggttcctgag ctgggagagc tggaacctga aagtgaaaga ggaatacgag 780

aaggtcgaga aagagtacaa gaccctggaa gagaggatca aagaggacat ccaggctctg 840

aaggctctgg aacagtatga gaaagagcgg caagaacagc tgctgcggga caccctgaac 900

accaacgagt accggctgag caagagaggc cttagaggct ggcgggaaat catccagaaa 960

tggctgaaaa tggacgagaa cgagccctcc gagaagtacc tggaagtgtt caaggactac 1020

cagcggaagc accctagaga ggccggcgat tacagcgtgt acgagttcct gtccaagaaa 1080

gagaaccact tcatctggcg gaatcaccct gagtacccct acctgtacgc caccttctgc 1140

gagatcgaca agaaaaagaa ggacgccaag cagcaggcca ccttcacact ggccgatcct 1200

atcaatcacc ctctgtgggt ccgattcgag gaaagaagcg gcagcaacct gaacaagtac 1260

agaatcctga ccgagcagct gcacaccgag aagctgaaga aaaagctgac agtgcagctg 1320

gaccggctga tctaccctac agaatctggc ggctgggaag agaagggcaa agtggacatt 1380

gtgctgctgc ccagccggca gttctacaac cagatcttcc tggacatcga ggaaaagggc 1440

aagcacgcct tcacctacaa ggatgagagc atcaagttcc ctctgaaggg cacactcggc 1500

ggagccagag tgcagttcga cagagatcac ctgagaagat accctcacaa ggtggaaagc 1560

ggcaacgtgg gcagaatcta cttcaacatg accgtgaaca tcgagcctac agagtcccca 1620

gtgtccaagt ctctgaagat ccaccgggac gacttcccca aggtggtcaa cttcaagccc 1680

aaagaactga ccgagtggat caaggacagc aagggcaaga aactgaagtc cggcatcgag 1740

tccctggaaa tcggcctgag agtgatgagc atcgacctgg gacagagaca ggccgctgcc 1800

gcctctattt tcgaggtggt ggatcagaag cccgacatcg aaggcaagct gtttttccca 1860

atcaagggca ccgagctgta tgccgtgcac agagccagct tcaacatcaa gctgcccggc 1920

gagacactgg tcaagagcag agaagtgctg cggaaggcca gagaggacaa tctgaaactg 1980

atgaaccaga agctcaactt cctgcggaac gtgctgcact tccagcagtt cgaggacatc 2040

accgagagag agaagcgggt caccaagtgg atcagcagac aagagaacag cgacgtgccc 2100

ctggtgtacc aggatgagct gatccagatc cgcgagctga tgtacaagcc ttacaaggac 2160

tgggtcgcct tcctgaagca gctccacaag agactggaag tcgagatcgg caaagaagtg 2220

aagcactggc ggaagtccct gagcgacgga agaaagggcc tgtacggcat ctccctgaag 2280

aacatcgacg agatcgatcg gacccggaag ttcctgctga gatggtccct gaggcctacc 2340

gaacctggcg aagtgcgtag actggaaccc ggccagagat tcgccatcga ccagctgaat 2400

cacctgaacg ccctgaaaga agatcggctg aagaagatgg ccaacaccat catcatgcac 2460

gccctgggct actgctacga cgtgcggaag aagaaatggc aggctaagaa ccccgcctgc 2520

cagatcatcc tgttcgagga tctgagcaac tacaacccct acgaggaaag gtcccgcttc 2580

gagaacagca agctcatgaa gtggtccaga cgcgagatcc ccagacaggt tgcactgcag 2640

ggcgagatct atggcctgca agtgggagaa gtgggcgctc agttcagcag cagattccac 2700

gccaagacag gcagccctgg catcagatgt agcgtcgtga ccaaagagaa gctgcaggac 2760

aatcggttct tcaagaatct gcagagagag ggcagactga ccctggacaa aatcgccgtg 2820

ctgaaagagg gcgatctgta cccagacaaa ggcggcgaga agttcatcag cctgagcaag 2880

gatcggaagt gcgtgaccac acacgccgac atcaacgccg ctcagaacct gcagaagcgg 2940

ttctggacaa gaacccacgg cttctacaag gtgtactgca aggcctacca ggtggacggc 3000

cagaccgtgt acatccctga gagcaaggac cagaagcaga agatcatcga agagttcggc 3060

gagggctact tcattctgaa ggacggggtg tacgaatggg tcaacgccgg caagggaggc 3120

tctggaggaa gctccgaagt cgagttttcc catgagtact ggatgagaca cgcattgact 3180

ctcgcaaaga gggctcgaga tgaacgcgag gtgcccgtgg gggcagtact cgtgctcaac 3240

aatcgcgtaa tcggcgaagg ttggaatagg gcaatcggac tccacgaccc cactgcacat 3300

gcggaaatca tggcccttcg acagggaggg cttgtgatgc agaattatcg actttatgat 3360

gcgacgctgt acgtcacgtt tgaaccttgc gtaatgtgcg cgggagctat gattcactcc 3420

cgcattggac gagttgtatt cggtgttcgc aacgccaaga cgggtgccgc aggttcactg 3480

atggacgtgc tgcatcatcc aggcatgaac caccgggtag aaatcacaga aggcatattg 3540

gcggacgaat gtgcggcgct gttgtgtcgt ttttttcgca tgcccaggcg ggtctttaac 3600

gcccagaaaa aagcacaatc ctctactgac ggctcttctg gatctgaaac acctggcaca 3660

agcgagagcg ccacccctga gagctctggc ctgaaaatca agaagggcag ctccaagcag 3720

agcagcagcg agctggtgga tagcgacatc ctgaaagaca gcttcgacct ggcctccgag 3780

ctgaaaggcg aaaagctgat gctgtacagg gaccccagcg gcaatgtgtt ccccagcgac 3840

aaatggatgg ccgctggcgt gttcttcgga aagctggaac gcatcctgat cagcaagctg 3900

accaaccagt actccatcag caccatcgag gacgacagca gcaagcagtc tatgaaaagg 3960

ccggcggcca cgaaaaaggc cggccaggca aaaaagaaaa agggatccta cccatacgat 4020

gttccagatt acgcttatcc ctacgacgtg cctgattatg catacccata tgatgtcccc 4080

gactatgcct aa 4092

<210> 11
<211> 1361
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: BhCas12b GGSGGS-ABE8-Xten20 at K1019 polypeptide sequence

<400> 11

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
20 25 30

Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
35 40 45

Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
50 55 60

```
His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65                  70                  75                  80

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
                85                  90                  95

Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
            100                 105                 110

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
        115                 120                 125

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
    130                 135                 140

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145                 150                 155                 160

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
                165                 170                 175

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
            180                 185                 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
        195                 200                 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
    210                 215                 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
225                 230                 235                 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
                245                 250                 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
            260                 265                 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
```

```
        275                 280                 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
    290                 295                 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
305                 310                 315                 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
                325                 330                 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
            340                 345                 350

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
        355                 360                 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
    370                 375                 380

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
385                 390                 395                 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
                405                 410                 415

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
            420                 425                 430

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
        435                 440                 445

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
    450                 455                 460

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

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
                485                 490                 495
```

```
Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
            500                 505                 510

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
        515                 520                 525

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
    530                 535                 540

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
545                 550                 555                 560

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
                565                 570                 575

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
            580                 585                 590

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
        595                 600                 605

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
    610                 615                 620

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
625                 630                 635                 640

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
                645                 650                 655

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
            660                 665                 670

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
        675                 680                 685

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
    690                 695                 700

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
```

```
705                 710                 715                 720

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
                725                 730                 735

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
            740                 745                 750

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
        755                 760                 765

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
    770                 775                 780

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
785                 790                 795                 800

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
                805                 810                 815

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
            820                 825                 830

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
        835                 840                 845

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
    850                 855                 860

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
865                 870                 875                 880

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
                885                 890                 895

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
            900                 905                 910

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
        915                 920                 925
```

```
Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
    930                 935                 940

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
945                 950                 955                 960

Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
                965                 970                 975

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
            980                 985                 990

Ala Tyr Gln Val Asp Gly Gln Thr  Val Tyr Ile Pro Glu  Ser Lys Asp
        995                 1000                1005

Gln Lys  Gln Lys Ile Ile Glu  Glu Phe Gly Glu Gly  Tyr Phe Ile
    1010                1015                1020

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

Ser Gly  Gly Ser Ser Glu Val  Glu Phe Ser His Glu  Tyr Trp Met
    1040                1045                1050

Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu
    1055                1060                1065

Val Pro  Val Gly Ala Val Leu  Val Leu Asn Asn Arg  Val Ile Gly
    1070                1075                1080

Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His Asp Pro  Thr Ala His
    1085                1090                1095

Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn
    1100                1105                1110

Tyr Arg  Leu Tyr Asp Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys
    1115                1120                1125

Val Met  Cys Ala Gly Ala Met  Ile His Ser Arg Ile  Gly Arg Val
```

```
    1130                1135                1140

Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu
    1145                1150                1155

Met Asp  Val Leu His His Pro  Gly Met Asn His Arg  Val Glu Ile
    1160                1165                1170

Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Arg
    1175                1180                1185

Phe Phe  Arg Met Pro Arg Arg  Val Phe Asn Ala Gln  Lys Lys Ala
    1190                1195                1200

Gln Ser  Ser Thr Asp Gly Ser  Ser Gly Ser Glu Thr  Pro Gly Thr
    1205                1210                1215

Ser Glu  Ser Ala Thr Pro Glu  Ser Ser Gly Leu Lys  Ile Lys Lys
    1220                1225                1230

Gly Ser  Ser Lys Gln Ser Ser  Ser Glu Leu Val Asp  Ser Asp Ile
    1235                1240                1245

Leu Lys  Asp Ser Phe Asp Leu  Ala Ser Glu Leu Lys  Gly Glu Lys
    1250                1255                1260

Leu Met  Leu Tyr Arg Asp Pro  Ser Gly Asn Val Phe  Pro Ser Asp
    1265                1270                1275

Lys Trp  Met Ala Ala Gly Val  Phe Phe Gly Lys Leu  Glu Arg Ile
    1280                1285                1290

Leu Ile  Ser Lys Leu Thr Asn  Gln Tyr Ser Ile Ser  Thr Ile Glu
    1295                1300                1305

Asp Asp  Ser Ser Lys Gln Ser  Met Lys Arg Pro Ala  Ala Thr Lys
    1310                1315                1320

Lys Ala  Gly Gln Ala Lys Lys  Lys Lys Gly Ser Tyr  Pro Tyr Asp
    1325                1330                1335
```

Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
    1340   1345   1350

Pro Tyr Asp Val Pro Asp Tyr Ala
    1355   1360

<210> 12
<211> 208
<212> PRT
<213> Petromyzon marinus

<220>
<221> MISC_FEATURE
<222> (1)..(208)
<223> tr|A5H718|A5H718_PETMA Cytosine deaminase OS=Petromyzon marinus
       OX=7757 PE=2 SV=1 amino acid sequence; PmCDA1 polypeptide
       sequence

<400> 12

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1   5   10   15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
            20   25   30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
        35   40   45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
    50   55   60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65   70   75   80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
                85   90   95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
            100   105   110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
        115   120   125

```
Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn
    130                 135                 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145                 150                 155                 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
                165                 170                 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
            180                 185                 190

Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
        195                 200                 205
```

<210> 13
<211> 766
<212> DNA
<213> Petromyzon marinus

<220>
<221> misc_feature
<222> (1)..(766)
<223> EF094822.1 Petromyzon marinus isolate PmCDA.21 cytosine deaminase mRNA, complete cds; PmCDA1 polynucleotide sequence

<400> 13

```
tgacacgaca cagccgtgta tatgaggaag ggtagctgga tgggggggggg gggaatacgt       60

tcagagagga cattagcgag cgtcttgttg gtggccttga gtctagacac ctgcagacat      120

gaccgacgct gagtacgtga gaatccatga gaagttggac atctacacgt ttaagaaaca      180

gtttttcaac aacaaaaaat ccgtgtcgca tagatgctac gttctctttg aattaaaacg      240

acggggtgaa cgtagagcgt gtttttgggg ctatgctgtg aataaaccac agagcgggac      300

agaacgtgga attcacgccg aaatctttag cattagaaaa gtcgaagaat acctgcgcga      360

caaccccgga caattcacga taaattggta ctcatcctgg agtccttgtg cagattgcgc      420

tgaaaagatc ttagaatggt ataaccagga gctgcggggg aacggccaca ctttgaaaat      480

ctgggcttgc aaactctatt acgagaaaaa tgcgaggaat caaattgggc tgtggaacct      540

cagagataac ggggttgggt tgaatgtaat ggtaagtgaa cactaccaat gttgcaggaa      600
```

aatattcatc caatcgtcgc acaatcaatt gaatgagaat agatggcttg agaagacttt 660

gaagcgagct gaaaaacgac ggagcgagtt gtccattatg attcaggtaa aaatactcca 720

caccactaag agtcctgctg tttaagaggc tatgcggatg gttttc 766

<210> 14
<211> 145
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(145)
<223> tr|Q6QJ80|Q6QJ80_HUMAN Activation-induced cytidine deaminase OS=Homo sapiens OX=9606 GN=AICDA PE=2 SV=1; AID polypeptide sequence

<400> 14

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 Ala Pro
130 135 140

Val
145

<210> 15
<211> 10681
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(10681)
<223> NG_011588.1:5001-15681 Homo sapiens activation induced cytidine deaminase (AICDA), RefSeqGene (LRG_17) on chromosome 12; nucleic acid sequence of the CDS of human AID; polynucleotide sequence

<400> 15
agagaaccat cattaattga agtgagattt ttctggcctg agacttgcag ggaggcaaga 60

agacactctg gacaccacta tggacaggta aagaggcagt cttctcgtgg gtgattgcac 120

tggccttcct ctcagagcaa atctgagtaa tgagactggt agctatccct ttctctcatg 180

taactgtctg actgataaga tcagcttgat caatatgcat atatattttt tgatctgtct 240

ccttttcttc tattcagatc ttatacgctg tcagcccaat tctttctgtt tcagacttct 300

cttgatttcc ctctttttca tgtggcaaaa gaagtagtgc gtacaatgta ctgattcgtc 360

ctgagatttg taccatggtt gaaactaatt tatggtaata atattaacat agcaaatctt 420

tagagactca aatcatgaaa aggtaatagc agtactgtac taaaaacggt agtgctaatt 480

ttcgtaataa ttttgtaaat attcaacagt aaaacaactt gaagacacac tttcctaggg 540

aggcgttact gaaataattt agctatagta agaaaatttg taattttaga aatgccaagc 600

attctaaatt aattgcttga aagtcactat gattgtgtcc attataagga gacaaattca 660

ttcaagcaag ttatttaatg ttaaaggccc aattgttagg cagttaatgg cacttttact 720

attaactaat ctttccattt gttcagacgt agcttaactt acctcttagg tgtgaatttg 780

gttaaggtcc tcataatgtc tttatgtgca gtttttgata ggttattgtc atagaactta 840

ttctattcct acatttatga ttactatgga tgtatgagaa taacacctaa tccttatact 900

ttacctcaat ttaactcctt tataaagaac ttacattaca gaataaagat tttttaaaaa 960

tatatttttt tgtagagaca gggtcttagc ccagccgagg ctggtctcta agtcctggcc 1020

caagcgatcc tcctgcctgg gcctcctaaa gtgctggaat tatagacatg agccatcaca 1080

tccaatatac agaataaaga tttttaatgg aggatttaat gttcttcaga aaattttctt 1140

gaggtcagac aatgtcaaat gtctcctcag tttacactga gattttgaaa acaagtctga 1200

gctataggtc cttgtgaagg gtccattgga aatacttgtt caaagtaaaa tggaaagcaa 1260

aggtaaaatc agcagttgaa attcagagaa agacagaaaa ggagaaaaga tgaaattcaa 1320

caggacagaa gggaaatata ttatcattaa ggaggacagt atctgtagag ctcattagtg 1380

atggcaaaat gacttggtca ggattatttt taacccgctt gtttctggtt tgcacggctg 1440

gggatgcagc tagggttctg cctcagggag cacagctgtc cagagcagct gtcagcctgc 1500

aagcctgaaa cactccctcg gtaaagtcct tcctactcag gacagaaatg acgagaacag 1560

ggagctggaa acaggcccct aaccagagaa gggaagtaat ggatcaacaa agttaactag 1620

caggtcagga tcacgcaatt catttcactc tgactggtaa catgtgacag aaacagtgta 1680

ggcttattgt attttcatgt agagtaggac ccaaaaatcc acccaaagtc ctttatctat 1740

gccacatcct tcttatctat acttccagga cacttttttct tccttatgat aaggctctct 1800

ctctctccac acacacacac acacacacac acacacacac acacacacac acaaacacac 1860

accccgccaa ccaaggtgca tgtaaaaaga tgtagattcc tctgcctttc tcatctacac 1920

agcccaggag ggtaagttaa tataagaggg atttattggt aagagatgat gcttaatctg 1980

tttaacactg ggcctcaaag agagaatttc ttttcttctg tacttattaa gcacctatta 2040

tgtgttgagc ttatatatac aaagggttat tatatgctaa tatagtaata gtaatggtgg 2100

ttggtactat ggtaattacc ataaaaatta ttatcctttt aaaataaagc taattattat 2160

tggatctttt ttagtattca ttttatgttt tttatgtttt tgatttttta aaagacaatc 2220

tcaccctgtt acccaggctg gagtgcagtg gtgcaatcat agctttctgc agtcttgaac 2280

tcctgggctc aagcaatcct cctgccttgg cctcccaaag tgttgggata cagtcatgag 2340

ccactgcatc tggcctagga tccatttaga ttaaaatatg cattttaaat tttaaaataa 2400

tatggctaat ttttacctta tgtaatgtgt atactggcaa taaatctagt ttgctgccta 2460

aagtttaaag tgctttccag taagcttcat gtacgtgagg ggagacattt aaagtgaaac 2520

agacagccag gtgtggtggc tcacgcctgt aatcccagca ctctgggagg ctgaggtggg 2580

tggatcgctt gagccctgga gttcaagacc agcctgagca acatggcaaa acgctgtttc 2640

tataacaaaa attagccggg catggtggca tgtgcctgtg gtcccagcta ctaggggct 2700

gaggcaggag aatcgttgga gcccaggagg tcaaggctgc actgagcagt gcttgcgcca 2760

ctgcactcca gcctgggtga caggaccaga ccttgcctca aaaaaataag aagaaaaatt 2820

aaaaataaat ggaaacaact acaaagagct gttgtcctag atgagctact tagttaggct 2880

gatattttgg tatttaactt ttaaagtcag ggtctgtcac ctgcactaca ttattaaaat 2940

atcaattctc aatgtatatc cacacaaaga ctggtacgtg aatgttcata gtacctttat 3000

tcacaaaacc ccaaagtaga gactatccaa atatccatca acaagtgaac aaataaacaa 3060

aatgtgctat atccatgcaa tggaatacca ccctgcagta caaagaagct acttggggat 3120

gaatcccaaa gtcatgacgc taaatgaaag agtcagacat gaaggaggag ataatgtatg 3180

ccatacgaaa ttctagaaaa tgaaagtaac ttatagttac agaaagcaaa tcagggcagg 3240

catagaggct cacacctgta atcccagcac tttgagaggc cacgtgggaa gattgctaga 3300

actcaggagt tcaagaccag cctgggcaac acagtgaaac tccattctcc acaaaaatgg 3360

gaaaaaaaga aagcaaatca gtggttgtcc tgtggggagg ggaaggactg caaagaggga 3420

agaagctctg gtggggtgag ggtggtgatt caggttctgt atcctgactg tggtagcagt 3480

ttggggtgtt tacatccaaa aatattcgta gaattatgca tcttaaatgg gtggagttta 3540

ctgtatgtaa attatacctc aatgtaagaa aaaataatgt gtaagaaaac tttcaattct 3600

cttgccagca aacgttattc aaattcctga gccctttact tcgcaaattc tctgcacttc 3660

tgccccgtac cattaggtga cagcactagc tccacaaatt ggataaatgc atttctggaa 3720

aagactaggg acaaaatcca ggcatcactt gtgctttcat atcaaccatg ctgtacagct 3780

tgtgttgctg tctgcagctg caatggggac tcttgatttc tttaaggaaa cttgggttac 3840

cagagtattt ccacaaatgc tattcaaatt agtgcttatg atatgcaaga cactgtgcta 3900

ggagccagaa aacaaagagg aggagaaatc agtcattatg tgggaacaac atagcaagat 3960

atttagatca ttttgactag ttaaaaaagc agcagagtac aaaatcacac atgcaatcag 4020

tataatccaa atcatgtaaa tatgtgcctg tagaaagact agaggaataa acacaagaat 4080

cttaacagtc attgtcatta gacactaagt ctaattatta ttattagaca ctatgatatt 4140

tgagatttaa aaaatcttta atattttaaa atttagagct cttctatttt tccatagtat 4200

tcaagtttga caatgatcaa gtattactct ttctttttt ttttttttt ttttttttga 4260

gatggagttt tggtcttgtt gcccatgctg gagtggaatg gcatgaccat agctcactgc 4320

aacctccacc tcctgggttc aagcaaagct gtcgcctcag cctcccgggt agatgggatt 4380

acaggcgccc accaccacac tcggctaatg tttgtatttt tagtagagat ggggtttcac 4440

catgttggcc aggctggtct caaactcctg acctcagagg atccacctgc ctcagcctcc 4500

caaagtgctg ggattacaga tgtaggccac tgcgcccggc caagtattgc tcttatacat 4560

taaaaaacag gtgtgagcca ctgcgcccag ccaggtattg ctcttataca ttaaaaaata 4620

ggccggtgca gtggctcacg cctgtaatcc cagcactttg ggaagccaag gcgggcagaa 4680

cacccgaggt caggagtcca aggccagcct ggccaagatg gtgaaacccc gtctctatta 4740

aaaatacaaa cattacctgg gcatgatggt gggcgcctgt aatcccagct actcaggagg 4800

ctgaggcagg aggatccgcg gagcctggca gatctgcctg agcctgggag gttgaggcta 4860

cagtaagcca agatcatgcc agtatacttc agcctgggcg acaaagtgag accgtaacaa 4920

aaaaaaaaaa atttaaaaaa agaaatttag atcaagatcc aactgtaaaa agtggcctaa 4980

acaccacatt aaagagtttg gagtttattc tgcaggcaga agagaaccat caggggtct 5040

tcagcatggg aatggcatgg tgcacctggt ttttgtgaga tcatggtggt gacagtgtgg 5100

ggaatgttat tttggaggga ctggaggcag acagaccggt taaaaggcca gcacaacaga 5160

taaggaggaa gaagatgagg gcttggaccg aagcagagaa gagcaaacag ggaaggtaca 5220

aattcaagaa atattggggg gtttgaatca acacatttag atgattaatt aaatatgagg 5280

actgaggaat aagaaatgag tcaaggatgg ttccaggctg ctaggctgct tacctgaggt 5340

ggcaaagtcg ggaggagtgg cagtttagga caggggggcag ttgaggaata ttgttttgat 5400

cattttgagt ttgaggtaca agttggacac ttaggtaaag actggagggg aaatctgaat 5460

atacaattat gggactgagg aacaagttta ttttattttt tgtttcgttt tcttgttgaa 5520

gaacaaattt aattgtaatc ccaagtcatc agcatctaga agacagtggc aggaggtgac 5580

tgtcttgtgg gtaagggttt ggggtccttg atgagtatct ctcaattggc cttaaatata 5640

agcaggaaaa ggagtttatg atggattcca ggctcagcag ggctcaggag ggctcaggca 5700

gccagcagag gaagtcagag catcttcttt ggtttagccc aagtaatgac ttccttaaaa 5760

agctgaagga aaatccagag tgaccagatt ataaactgta ctcttgcatt ttctctccct 5820

cctctcaccc acagcctctt gatgaaccgg aggaagtttc tttaccaatt caaaaatgtc 5880

cgctgggcta agggtcggcg tgagacctac ctgtgctacg tagtgaagag gcgtgacagt 5940

gctacatcct tttcactgga ctttggttat cttcgcaata aggtatcaat taaagtcggc 6000

tttgcaagca gtttaatggt caactgtgag tgcttttaga gccacctgct gatggtatta 6060

cttccatcct tttttggcat ttgtgtctct atcacattcc tcaaatcctt ttttttattt 6120

ctttttccat gtccatgcac ccatattaga catggcccaa aatatgtgat ttaattcctc 6180

cccagtaatg ctgggcaccc taataccact ccttccttca gtgccaagaa caactgctcc 6240

caaactgttt accagctttc ctcagcatct gaattgcctt tgagattaat taagctaaaa 6300

gcatttttat atgggagaat attatcagct tgtccaagca aaaattttaa atgtgaaaaa 6360

caaattgtgt cttaagcatt tttgaaaatt aaggaagaag aatttgggaa aaaattaacg 6420

gtggctcaat tctgtcttcc aaatgatttc ttttccctcc tactcacatg ggtcgtaggc 6480

cagtgaatac attcaacatg gtgatcccca gaaaactcag agaagcctcg gctgatgatt 6540

aattaaattg atctttcggc tacccgagag aattacattt ccaagagact tcttcaccaa 6600

aatccagatg ggtttacata aacttctgcc cacgggtatc tcctctctcc taacacgctg 6660

tgacgtctgg gcttggtgga atctcaggga agcatccgtg gggtggaagg tcatcgtctg 6720

gctcgttgtt tgatggttat attaccatgc aattttcttt gcctacattt gtattgaata 6780

catcccaatc tccttcctat tcggtgacat gacacattct atttcagaag gctttgattt 6840

tatcaagcac tttcatttac ttctcatggc agtgcctatt acttctctta caatacccat 6900

ctgtctgctt taccaaaatc tatttcccct tttcagatcc tcccaaatgg tcctcataaa 6960

ctgtcctgcc tccacctagt ggtccaggta tatttccaca atgttacatc aacaggcact 7020

tctagccatt ttccttctca aaaggtgcaa aaagcaactt cataaacaca aattaaatct 7080

tcggtgaggt agtgtgatgc tgcttcctcc caactcagcg cacttcgtct tcctcattcc 7140

acaaaaaccc atagccttcc ttcactctgc aggactagtg ctgccaaggg ttcagctcta 7200

cctactggtg tgctcttttg agcaagttgc ttagcctctc tgtaacacaa ggacaatagc 7260

tgcaagcatc cccaaagatc attgcaggag acaatgacta aggctaccag agccgcaata 7320

aaagtcagtg aattttagcg tggtcctctc tgtctctcca gaacggctgc cacgtggaat 7380

tgctcttcct ccgctacatc tcggactggg acctagaccc tggccgctgc taccgcgtca 7440

cctggttcac ctcctggagc ccctgctacg actgtgcccg acatgtggcc gactttctgc 7500

gagggaaccc caacctcagt ctgaggatct tcaccgcgcg cctctacttc tgtgaggacc 7560

gcaaggctga gcccgagggg ctgcggcggc tgcaccgcgc cggggtgcaa atagccatca 7620

tgaccttcaa aggtgcgaaa gggccttccg cgcaggcgca gtgcagcagc ccgcattcgg 7680

gattgcgatg cggaatgaat gagttagtgg ggaagctcga ggggaagaag tgggcgggga 7740

ttctggttca cctctggagc cgaaattaaa gattagaagc agagaaaaga gtgaatggct 7800

cagagacaag gccccgagga aatgagaaaa tggggccagg gttgcttctt tcccctcgat 7860

ttggaacctg aactgtcttc tacccccata tccccgcctt tttttccttt ttttttttt 7920

gaagattatt tttactgctg gaatactttt gtagaaaacc acgaaagaac tttcaaagcc 7980

tgggaagggc tgcatgaaaa ttcagttcgt ctctccagac agcttcggcg catccttttg 8040

gtaaggggct tcctcgcttt ttaaattttc tttctttctc tacagtcttt tttggagttt 8100

cgtatatttc ttatattttc ttattgttca atcactctca gttttcatct gatgaaaact 8160

ttatttctcc tccacatcag ctttttcttc tgctgtttca ccattcagag ccctctgcta 8220

aggttccttt tccctccctt ttctttcttt tgttgtttca catctttaaa tttctgtctc 8280

tccccagggt tgcgtttcct tcctggtcag aattcttttc tccttttttt ttttttttt 8340

ttttttttt aaacaaacaa acaaaaaacc caaaaaaact ctttcccaat ttactttctt 8400

ccaacatgtt acaaagccat ccactcagtt tagaagactc tccggcccca ccgaccccca 8460

acctcgtttt gaagccattc actcaatttg cttctctctt tctctacagc ccctgtatga 8520

ggttgatgac ttacgagacg catttcgtac tttgggactt tgatagcaac ttccaggaat 8580

gtcacacacg atgaaatatc tctgctgaag acagtggata aaaaacagtc cttcaagtct 8640

tctctgtttt tattcttcaa ctctcacttt cttagagttt acagaaaaaa tatttatata 8700

cgactcttta aaaagatcta tgtcttgaaa atagagaagg aacacaggtc tggccaggga 8760

cgtgctgcaa ttggtgcagt tttgaatgca acattgtccc ctactgggaa taacagaact 8820

gcaggacctg ggagcatcct aaagtgtcaa cgtttttcta tgacttttag gtaggatgag 8880

agcagaaggt agatcctaaa aagcatggtg agaggatcaa atgtttttat atcaacatcc 8940

tttattattt gattcatttg agttaacagt ggtgttagtg atagattttt ctattctttt 9000

cccttgacgt ttactttcaa gtaacacaaa ctcttccatc aggccatgat ctataggacc 9060

tcctaatgag agtatctggg tgattgtgac cccaaaccat ctctccaaag cattaatatc 9120

caatcatgcg ctgtatgttt taatcagcag aagcatgttt ttatgtttgt acaaaagaag 9180

attgttatgg gtggggatgg aggtatagac catgcatggt caccttcaag ctactttaat 9240

aaaggatctt aaaatgggca ggaggactgt gaacaagaca ccctaataat gggttgatgt 9300

ctgaagtagc aaatcttctg gaaacgcaaa ctcttttaag gaagtcccta atttagaaac 9360

acccacaaac ttcacatatc ataattagca aacaattgga aggaagttgc ttgaatgttg 9420

gggagaggaa aatctattgg ctctcgtggg tctcttcatc tcagaaatgc caatcaggtc 9480

aaggtttgct acattttgta tgtgtgtgat gcttctccca aaggtatatt aactatataa 9540

gagagttgtg acaaaacaga atgataaagc tgcgaaccgt ggcacacgct catagttcta 9600

gctgcttggg aggttgagga gggaggatgg cttgaacaca ggtgttcaag gccagcctgg 9660

gcaacataac aagatcctgt ctctcaaaaa aaaaaaaaaa aaaaagaaag agagagggcc 9720

gggcgtggtg gctcacgcct gtaatcccag cactttggga ggccgagccg ggcggatcac 9780

ctgtggtcag gagtttgaga ccagcctggc caacatggca aaaccccgtc tgtactcaaa 9840

atgcaaaaat tagccaggcg tggtagcagg cacctgtaat cccagctact tgggaggctg 9900

aggcaggaga atcgcttgaa cccaggaggt ggaggttgca gtaagctgag atcgtgccgt 9960

tgcactccag cctgggcgac aagagcaaga ctctgtctca gaaaaaaaaa aaaaaaagag 10020

agagagagag aaagagaaca atatttggga gagaaggatg gggaagcatt gcaaggaaat 10080

tgtgctttat ccaacaaaat gtaaggagcc aataagggat ccctatttgt ctcttttggt 10140

gtctatttgt ccctaacaac tgtctttgac agtgagaaaa atattcagaa taaccatatc 10200

cctgtgccgt tattacctag caacccttgc aatgaagatg agcagatcca caggaaaact 10260

tgaatgcaca actgtcttat tttaatctta ttgtacataa gtttgtaaaa gagttaaaaa 10320

ttgttacttc atgtattcat ttatatttta tattattttg cgtctaatga ttttttatta 10380

acatgatttc cttttctgat atattgaaat ggagtctcaa agcttcataa atttataact 10440

ttagaaatga ttctaataac aacgtatgta attgtaacat tgcagtaatg gtgctacgaa 10500

gccatttctc ttgattttta gtaaactttt atgacagcaa atttgcttct ggctcacttt 10560

caatcagtta aataaatgat aaataatttt ggaagctgtg aagataaaat accaaataaa 10620

ataatataaa agtgatttat atgaagttaa aataaaaaat cagtatgatg gaataaactt 10680

<210> 16
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(198)
<223> human AID polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(30)
<223> Nuclear localization sequence

<220>
<221> MISC_FEATURE
<222> (183)..(198)
<223> Nuclear export signal

<400> 16

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> 17
<211> 198
<212> PRT
<213> Canis lupus familiaris

<220>
<221> MISC_FEATURE
<222> (1)..(198)
<223> Canine AID (clAID) polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(30)
<223> Nuclear localization sequence

<220>
<221> MISC_FEATURE
<222> (183)..(198)
<223> Nuclear export signal

<400> 17

Met Asp Ser Leu Leu Met Lys Gln Arg Lys Phe Leu Tyr His Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His 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 His
        35                  40                  45

Leu Arg Asn Lys Ser 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 Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala 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 Arg Glu Lys 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> 18
<211> 199
<212> PRT
<213> Bos Taurus

<220>
<221> MISC_FEATURE

<222> (1)..(199)
<223> Bovine AID (btAID) polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(30)
<223> Nuclear localization sequence

<220>
<221> MISC_FEATURE
<222> (183)..(199)
<223> Nuclear export signal

<400> 18

```
Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys
1               5                  10                  15


Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
            20                  25                  30


Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His
        35                  40                  45


Leu Arg Asn Lys Ala 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 Tyr Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
            100                 105                 110


Leu Tyr Phe Cys Asp Lys Glu Arg Lys Ala Glu Pro Glu Gly Leu Arg
        115                 120                 125


Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp
    130                 135                 140


Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe
145                 150                 155                 160
```

Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln
165 170 175

Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp
180 185 190

Ala Phe Arg Thr Leu Gly Leu
195

<210> 19
<211> 239
<212> PRT
<213> Rattus

<220>
<221> MISC_FEATURE
<222> (1)..(239)
<223> Rat AID polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(26)
<223> Nuclear localization sequence

<220>
<221> MISC_FEATURE
<222> (194)..(239)
<223> Nuclear export signal

<400> 19

Met Ala Val Gly Ser Lys Pro Lys Ala Ala Leu Val Gly Pro His Trp
1 5 10 15

Glu Arg Glu Arg Ile Trp Cys Phe Leu Cys Ser Thr Gly Leu Gly Thr
20 25 30

Gln Gln Thr Gly Gln Thr Ser Arg Trp Leu Arg Pro Ala Ala Thr Gln
35 40 45

Asp Pro Val Ser Pro Pro Arg Ser Leu Leu Met Lys Gln Arg Lys Phe
50 55 60

Leu Tyr His Phe Lys Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr
65 70 75 80

```
Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser
                85                  90                  95

Leu Asp Phe Gly Tyr Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu
            100                 105                 110

Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
        115                 120                 125

Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
    130                 135                 140

Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg
145                 150                 155                 160

Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu
                165                 170                 175

Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe
            180                 185                 190

Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu
        195                 200                 205

Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu
    210                 215                 220

Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly Leu
225                 230                 235
```

<210> 20
<211> 198
<212> PRT
<213> Mus musculus

<220>
<221> MISC_FEATURE
<222> (1)..(198)
<223> Mouse (mAID) AID polypeptide sequence

<400> 20

```
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> 21

<211> 229
<212> PRT
<213> Rattus norvegicus

<220>
<221> MISC_FEATURE
<222> (1)..(229)
<223> rAPOBEC-1 polypeptide sequence

<400> 21

```
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>  22
<211>  229
<212>  PRT
<213>  Mesocricetus auratus


<220>
<221>  MISC_FEATURE
<222>  (1)..(229)
<223>  maAPOBEC-1 polypeptide sequence

<400>  22

Met Ser Ser Glu Thr Gly Pro Val Val Val Asp Pro Thr Leu Arg Arg
1               5                   10                  15


Arg Ile Glu Pro His Glu Phe Asp Ala Phe Phe Asp Gln Gly Glu Leu
            20                  25                  30


Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Arg Trp Gly Gly Arg His
        35                  40                  45


Asn Ile Trp Arg His Thr Gly Gln Asn Thr Ser Arg His Val Glu Ile
    50                  55                  60


Asn Phe Ile Glu Lys Phe Thr Ser Glu Arg Tyr Phe Tyr Pro Ser Thr
65                  70                  75                  80


Arg Cys Ser Ile Val Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
                85                  90                  95
```

```
Ser Lys Ala Ile Thr Glu Phe Leu Ser Gly His Pro Asn Val Thr Leu
            100                 105                 110

Phe Ile Tyr Ala Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg
        115                 120                 125

Gln Gly Leu Arg Asp Leu Ile Ser Arg Gly Val Thr Ile Arg Ile Met
    130                 135                 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro
145                 150                 155                 160

Pro Ser Asn Glu Val Tyr Trp Pro Arg Tyr Pro Asn Leu Trp Met Arg
                165                 170                 175

Leu Tyr Ala Leu Glu Leu Tyr Cys Ile His Leu Gly Leu Pro Pro Cys
            180                 185                 190

Leu Lys Ile Lys Arg Arg His Gln Tyr Pro Leu Thr Phe Phe Arg Leu
        195                 200                 205

Asn Leu Gln Ser Cys His Tyr Gln Arg Ile Pro Pro His Ile Leu Trp
    210                 215                 220

Ala Thr Gly Phe Ile
225
```

<210> 23
<211> 236
<212> PRT
<213> Pongo pygmaeus

<220>
<221> MISC_FEATURE
<222> (1)..(236)
<223> ppAPOBEC-1 polypeptide sequence

<400> 23

```
Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
1               5                   10                  15
```

```
Arg Ile Glu Ser Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu
            20                  25                  30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
        35                  40                  45

Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
    50                  55                  60

Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Arg Phe His Ser Ser Ile
65                  70                  75                  80

Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
                85                  90                  95

Ser Gln Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu
            100                 105                 110

Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Arg Asn Arg
        115                 120                 125

Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
    130                 135                 140

Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145                 150                 155                 160

Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
                165                 170                 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
            180                 185                 190

Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Ala Phe Phe Arg Leu
        195                 200                 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
    210                 215                 220

Ala Thr Gly Leu Ile His Pro Ser Val Thr Trp Arg
225                 230                 235
```

<210> 24
<211> 236
<212> PRT
<213> Oryctolagus cuniculus

<220>
<221> MISC_FEATURE
<222> (1)..(236)
<223> ocAPOBEC1 polypeptide sequence

<400> 24

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

Arg Ile Glu Pro Trp Glu Phe Glu Val Phe Phe Asp Pro Gln Glu Leu
20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Ala Ser Ser
35 40 45

Lys Thr Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60

Asn Phe Leu Glu Lys Leu Thr Ser Glu Gly Arg Leu Gly Pro Ser Thr
65 70 75 80

Cys Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95

Ser Met Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu
100 105 110

Ile Ile Phe Val Ala Arg Leu Phe Gln His Met Asp Arg Arg Asn Arg
115 120 125

Gln Gly Leu Lys Asp Leu Val Thr Ser Gly Val Thr Val Arg Val Met
130 135 140

Ser Val Ser Glu Tyr Cys Tyr Cys Trp Glu Asn Phe Val Asn Tyr Pro
145 150 155 160

```
Pro Gly Lys Ala Ala Gln Trp Pro Arg Tyr Pro Pro Arg Trp Met Leu
                165                 170                 175

Met Tyr Ala Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
            180                 185                 190

Leu Lys Ile Ser Arg Arg His Gln Lys Gln Leu Thr Phe Phe Ser Leu
        195                 200                 205

Thr Pro Gln Tyr Cys His Tyr Lys Met Ile Pro Pro Tyr Ile Leu Leu
    210                 215                 220

Ala Thr Gly Leu Leu Gln Pro Ser Val Pro Trp Arg
225                 230                 235
```

<210> 25
<211> 235
<212> PRT
<213> Monodelphis domestica

<220>
<221> MISC_FEATURE
<222> (1)..(235)
<223> mdAPOBEC-1 polypeptide sequence

<400> 25

```
Met Asn Ser Lys Thr Gly Pro Ser Val Gly Asp Ala Thr Leu Arg Arg
1               5                   10                  15

Arg Ile Lys Pro Trp Glu Phe Val Ala Phe Phe Asn Pro Gln Glu Leu
            20                  25                  30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Asn Gln Asn
        35                  40                  45

Ile Trp Arg His Ser Asn Gln Asn Thr Ser Gln His Ala Glu Ile Asn
    50                  55                  60

Phe Met Glu Lys Phe Thr Ala Glu Arg His Phe Asn Ser Ser Val Arg
65                  70                  75                  80
```

```
Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser
                85                  90                  95

Lys Ala Ile Arg Lys Phe Leu Asp His Tyr Pro Asn Val Thr Leu Ala
            100                 105                 110

Ile Phe Ile Ser Arg Leu Tyr Trp His Met Asp Gln Gln His Arg Gln
        115                 120                 125

Gly Leu Lys Glu Leu Val His Ser Gly Val Thr Ile Gln Ile Met Ser
    130                 135                 140

Tyr Ser Glu Tyr His Tyr Cys Trp Arg Asn Phe Val Asp Tyr Pro Gln
145                 150                 155                 160

Gly Glu Glu Asp Tyr Trp Pro Lys Tyr Pro Tyr Leu Trp Ile Met Leu
                165                 170                 175

Tyr Val Leu Glu Leu His Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu
            180                 185                 190

Lys Ile Ser Gly Ser His Ser Asn Gln Leu Ala Leu Phe Ser Leu Asp
        195                 200                 205

Leu Gln Asp Cys His Tyr Gln Lys Ile Pro Tyr Asn Val Leu Val Ala
    210                 215                 220

Thr Gly Leu Val Gln Pro Phe Val Thr Trp Arg
225                 230                 235
```

<210> 26
<211> 224
<212> PRT
<213> Pongo pygmaeus

<220>
<221> MISC_FEATURE
<222> (1)..(224)
<223> ppAPOBEC-2 polypeptide sequence

<400> 26

```
Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Thr Glu Ala Ala Ser Gln
```

```
1               5                   10                  15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
            20                  25                  30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
        35                  40                  45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
    50                  55                  60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65                  70                  75                  80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
                85                  90                  95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
            100                 105                 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
        115                 120                 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
    130                 135                 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Leu
145                 150                 155                 160

Glu Ile Gln Asp Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
                165                 170                 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
            180                 185                 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile
        195                 200                 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
    210                 215                 220
```

<210> 27
<211> 224
<212> PRT
<213> Bos Taurus

<220>
<221> MISC_FEATURE
<222> (1)..(224)
<223> btAPOBEC-2 polypeptide sequence

<400> 27

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln
1   5   10   15

Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
            20   25   30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
        35   40   45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
    50   55   60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
65   70   75   80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr
                85   90   95

Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp
            100   105   110

Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
        115   120   125

Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn Lys Thr Lys Asn
    130   135   140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145   150   155   160

Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 28
<211> 440
<212> PRT
<213> Mus musculus

<220>
<221> MISC_FEATURE
<222> (1)..(440)
<223> mAPOBEC-3-(1) polypeptide sequence

<400> 28

Met Gln Pro Gln Arg Leu Gly Pro Arg Ala Gly Met Gly Pro Phe Cys
1 5 10 15

Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro Ile Arg Asn Leu Ile
20 25 30

Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn Leu Gly Tyr Ala Lys
35 40 45

Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val Thr Arg Lys Asp Cys
50 55 60

Asp Ser Pro Val Ser Leu His His Gly Val Phe Lys Asn Lys Asp Asn
65 70 75 80

Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe His Asp Lys Val Leu
85 90 95

Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile Thr Trp Tyr Met Ser

```
            100                 105                 110

Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile Val Arg Phe Leu Ala
        115                 120                 125

Thr His His Asn Leu Ser Leu Asp Ile Phe Ser Ser Arg Leu Tyr Asn
    130                 135                 140

Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys Arg Leu Val Gln Glu
145                 150                 155                 160

Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu Phe Lys Lys Cys Trp
                165                 170                 175

Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe Arg Pro Trp Lys Arg
            180                 185                 190

Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys Leu Gln Glu Ile Leu
        195                 200                 205

Arg Pro Cys Tyr Ile Ser Val Pro Ser Ser Ser Ser Ser Thr Leu Ser
    210                 215                 220

Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr Arg Phe Trp Val Glu
225                 230                 235                 240

Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu Phe Tyr Ser Gln Phe
                245                 250                 255

Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr His Arg Met Lys Pro
            260                 265                 270

Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly Gln Ala Pro Leu Lys
        275                 280                 285

Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His Ala Glu Ile Leu Phe
    290                 295                 300

Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln Val Thr Ile Thr Cys
305                 310                 315                 320
```

```
Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala Trp Gln Leu Ala Ala
                325                 330                 335

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

Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys Gly Leu Cys Ser Leu
        355                 360                 365

Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp Leu Pro Gln Phe Thr
    370                 375                 380

Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg Pro Phe Trp Pro Trp
385                 390                 395                 400

Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln Arg Arg Leu Arg Arg
                405                 410                 415

Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val Asn Asp Phe Gly Asn
            420                 425                 430

Leu Gln Leu Gly Pro Pro Met Ser
        435                 440
```

<210> 29
<211> 429
<212> PRT
<213> Mus musculus

<220>
<221> MISC_FEATURE
<222> (1)..(429)
<223> APOBEC-3-(2) (Mouse APOBEC-3) polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (71)..(108)
<223> Nucleic acid editing domain

<220>
<221> MISC_FEATURE
<222> (288)..(319)
<223> Nucleic acid editing domain

<400> 29

```
Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1               5                  10                  15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
            20                  25                  30

Leu Gly Tyr Ala Lys Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val
        35                  40                  45

Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe
    50                  55                  60

Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe
65                  70                  75                  80

His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile
                85                  90                  95

Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile
            100                 105                 110

Val Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser
        115                 120                 125

Ser Arg Leu Tyr Asn Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys
    130                 135                 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu
145                 150                 155                 160

Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe
                165                 170                 175

Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys
            180                 185                 190

Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser
        195                 200                 205
```

```
Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr
    210                 215                 220

Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu
225                 230                 235                 240

Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr
                245                 250                 255

His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly
            260                 265                 270

Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His
        275                 280                 285

Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln
    290                 295                 300

Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
305                 310                 315                 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
                325                 330                 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
            340                 345                 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
        355                 360                 365

Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg
    370                 375                 380

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

Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val
                405                 410                 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
            420                 425
```

<210> 30
<211> 430
<212> PRT
<213> Rattus

<220>
<221> MISC_FEATURE
<222> (1)..(430)
<223> APOBEC-3 polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (72)..(109)
<223> Nucleic acid editing domain

<220>
<221> MISC_FEATURE
<222> (289)..(320)
<223> Nucleic acid editing domain

<400> 30

```
Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1               5                   10                  15


Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
            20                  25                  30


Arg Leu Arg Tyr Ala Ile Asp Arg Lys Asp Thr Phe Leu Cys Tyr Glu
        35                  40                  45


Val Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val
    50                  55                  60


Phe Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp
65                  70                  75                  80


Phe His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys
                85                  90                  95


Ile Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln
            100                 105                 110


Val Leu Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe
```

```
        115                 120                 125

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

Cys Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr
145                 150                 155                 160

Glu Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg
                165                 170                 175

Phe Arg Pro Trp Lys Lys Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser
            180                 185                 190

Lys Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser
        195                 200                 205

Ser Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu
    210                 215                 220

Thr Arg Phe Cys Val Glu Arg Arg Arg Val His Leu Leu Ser Glu Glu
225                 230                 235                 240

Glu Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr
                245                 250                 255

Tyr His Gly Val Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn
            260                 265                 270

Gly Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln
        275                 280                 285

His Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser
    290                 295                 300

Gln Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys
305                 310                 315                 320

Ala Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu
                325                 330                 335
```

```
His Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln
            340                 345                 350


Lys Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met
        355                 360                 365


Asp Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys
    370                 375                 380


Arg Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr
385                 390                 395                 400


Gln Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu
                405                 410                 415


Val Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
            420                 425                 430


<210>  31
<211>  199
<212>  PRT
<213>  Homo sapiens


<220>
<221>  MISC_FEATURE
<222>  (1)..(199)
<223>  hAPOBEC-3A polypeptide sequence

<220>
<221>  MISC_FEATURE
<222>  (70)..(106)
<223>  Nucleic acid editing domain

<400>  31

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>  32
<211>  373
<212>  PRT
<213>  Homo sapiens


<220>
<221>  MISC_FEATURE
<222>  (1)..(373)
<223>  hAPOBEC-3F polypeptide sequence

<220>
<221>  MISC_FEATURE
```

<222> (65)..(99)
<223> Nucleic acid editing domain

<220>
<221> MISC_FEATURE
<222> (248)..(283)
<223> Nucleic acid editing domain

<400> 32

```
Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1               5                   10                  15


Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
            20                  25                  30


Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg
        35                  40                  45


Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His
    50                  55                  60


His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro
65                  70                  75                  80


Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys
                85                  90                  95


Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn
            100                 105                 110


Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg
        115                 120                 125


Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val
    130                 135                 140


Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val
145                 150                 155                 160


Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn
                165                 170                 175
```

```
Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met
            180                 185                 190

Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg
        195                 200                 205

Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val
    210                 215                 220

Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn
225                 230                 235                 240

Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser
                245                 250                 255

Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr
            260                 265                 270

Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala
        275                 280                 285

Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala
    290                 295                 300

Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser
305                 310                 315                 320

Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe
                325                 330                 335

Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys
            340                 345                 350

Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu
        355                 360                 365

Gln Glu Ile Leu Glu
    370
```

<210> 33
<211> 370

<212> PRT
<213> Rhesus macaque

<220>
<221> MISC_FEATURE
<222> (1)..(370)
<223> Rhesus macaque APOBEC-3G polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(61)
<223> Cytoplasmic localization signal

<400> 33

Met Val Glu Pro Met Asp Pro Arg Thr Phe Val Ser Asn Phe Asn Asn
1 5 10 15

Arg Pro Ile Leu Ser Gly Leu Asn Thr Val Trp Leu Cys Cys Glu Val
20 25 30

Lys Thr Lys Asp Pro Ser Gly Pro Pro Leu Asp Ala Lys Ile Phe Gln
35 40 45

Gly Lys Val Tyr Ser Lys Ala Lys Tyr His Pro Glu Met Arg Phe Leu
50 55 60

Arg Trp Phe His Lys Trp Arg Gln Leu His His Asp Gln Glu Tyr Lys
65 70 75 80

Val Thr Trp Tyr Val Ser Trp Ser Pro Cys Thr Arg Cys Ala Asn Ser
85 90 95

Val Ala Thr Phe Leu Ala Lys Asp Pro Lys Val Thr Leu Thr Ile Phe
100 105 110

Val Ala Arg Leu Tyr Tyr Phe Trp Lys Pro Asp Tyr Gln Gln Ala Leu
115 120 125

Arg Ile Leu Cys Gln Lys Arg Gly Gly Pro His Ala Thr Met Lys Ile
130 135 140

Met Asn Tyr Asn Glu Phe Gln Asp Cys Trp Asn Lys Phe Val Asp Gly
145 150 155 160

```
Arg Gly Lys Pro Phe Lys Pro Arg Asn Asn Leu Pro Lys His Tyr Thr
                165                 170                 175

Leu Leu Gln Ala Thr Leu Gly Glu Leu Leu Arg His Leu Met Asp Pro
            180                 185                 190

Gly Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln
        195                 200                 205

His Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Leu His Asn Asp Thr
    210                 215                 220

Trp Val Pro Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Pro
225                 230                 235                 240

Asn Ile His Gly Phe Pro Lys Gly Arg His Ala Glu Leu Cys Phe Leu
                245                 250                 255

Asp Leu Ile Pro Phe Trp Lys Leu Asp Gly Gln Gln Tyr Arg Val Thr
            260                 265                 270

Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala
        275                 280                 285

Lys Phe Ile Ser Asn Asn Glu His Val Ser Leu Cys Ile Phe Ala Ala
    290                 295                 300

Arg Ile Tyr Asp Asp Gln Gly Arg Tyr Gln Glu Gly Leu Arg Ala Leu
305                 310                 315                 320

His Arg Asp Gly Ala Lys Ile Ala Met Met Asn Tyr Ser Glu Phe Glu
                325                 330                 335

Tyr Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe Gln Pro
            340                 345                 350

Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg
        355                 360                 365
```

Ala Ile
370

<210> 34
<211> 384
<212> PRT
<213> Pan troglodytes

<220>
<221> MISC_FEATURE
<222> (1)..(384)
<223> Chimpanzee APOBEC-3G polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(60)
<223> Cytoplasmic localization signal

<220>
<221> MISC_FEATURE
<222> (65)..(100)
<223> Nucleic acid editing domain

<220>
<221> MISC_FEATURE
<222> (257)..(291)
<223> Nucleic acid editing domain

<400> 34

Met Lys Pro His Phe Arg Asn Pro Val Glu Arg Met Tyr Gln Asp Thr
1 5 10 15

Phe Ser Asp Asn Phe Tyr Asn Arg Pro Ile Leu Ser His Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45

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

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95

```
Cys Thr Lys Cys Thr Arg Asp Val Ala Thr Phe Leu Ala Glu Asp Pro
            100                 105                 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
        115                 120                 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
    130                 135                 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145                 150                 155                 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
                165                 170                 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
            180                 185                 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Ser Asn Phe Asn Asn
        195                 200                 205

Glu Leu Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
    210                 215                 220

Glu Arg Leu His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225                 230                 235                 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
                245                 250                 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
            260                 265                 270

Leu His Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
        275                 280                 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Asn Asn Lys His
    290                 295                 300
```

```
Val Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305                 310                 315                 320


Cys Gln Glu Gly Leu Arg Thr Leu Ala Lys Ala Gly Ala Lys Ile Ser
                325                 330                 335


Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
            340                 345                 350


His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser
        355                 360                 365


Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
    370                 375                 380


<210>  35
<211>  377
<212>  PRT
<213>  Chlorocebus sabaeus


<220>
<221>  MISC_FEATURE
<222>  (1)..(377)
<223>  Green monkey APOBEC-3G polypeptide sequence

<220>
<221>  MISC_FEATURE
<222>  (1)..(60)
<223>  Cytoplasmic localization signal

<220>
<221>  MISC_FEATURE
<222>  (65)..(100)
<223>  Nucleic acid editing domain

<220>
<221>  MISC_FEATURE
<222>  (257)..(290)
<223>  Nucleic acid editing domain

<400>  35

Met Asn Pro Gln Ile Arg Asn Met Val Glu Gln Met Glu Pro Asp Ile
1               5                   10                  15


Phe Val Tyr Tyr Phe Asn Asn Arg Pro Ile Leu Ser Gly Arg Asn Thr
            20                  25                  30
```

```
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Asp Pro Ser Gly Pro Pro
        35                  40                  45

Leu Asp Ala Asn Ile Phe Gln Gly Lys Leu Tyr Pro Glu Ala Lys Asp
    50                  55                  60

His Pro Glu Met Lys Phe Leu His Trp Phe Arg Lys Trp Arg Gln Leu
65                  70                  75                  80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Val Ser Trp Ser Pro
                85                  90                  95

Cys Thr Arg Cys Ala Asn Ser Val Ala Thr Phe Leu Ala Glu Asp Pro
            100                 105                 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Lys
        115                 120                 125

Pro Asp Tyr Gln Gln Ala Leu Arg Ile Leu Cys Gln Glu Arg Gly Gly
    130                 135                 140

Pro His Ala Thr Met Lys Ile Met Asn Tyr Asn Glu Phe Gln His Cys
145                 150                 155                 160

Trp Asn Glu Phe Val Asp Gly Gln Gly Lys Pro Phe Lys Pro Arg Lys
                165                 170                 175

Asn Leu Pro Lys His Tyr Thr Leu Leu His Ala Thr Leu Gly Glu Leu
            180                 185                 190

Leu Arg His Val Met Asp Pro Gly Thr Phe Thr Ser Asn Phe Asn Asn
        195                 200                 205

Lys Pro Trp Val Ser Gly Gln Arg Glu Thr Tyr Leu Cys Tyr Lys Val
    210                 215                 220

Glu Arg Ser His Asn Asp Thr Trp Val Leu Leu Asn Gln His Arg Gly
225                 230                 235                 240
```

```
Phe Leu Arg Asn Gln Ala Pro Asp Arg His Gly Phe Pro Lys Gly Arg
                245                 250                 255

His Ala Glu Leu Cys Phe Leu Asp Leu Ile Pro Phe Trp Lys Leu Asp
            260                 265                 270

Asp Gln Gln Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
        275                 280                 285

Ser Cys Ala Gln Lys Met Ala Lys Phe Ile Ser Asn Asn Lys His Val
    290                 295                 300

Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys
305                 310                 315                 320

Gln Glu Gly Leu Arg Thr Leu His Arg Asp Gly Ala Lys Ile Ala Val
                325                 330                 335

Met Asn Tyr Ser Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Asp Arg
            340                 345                 350

Gln Gly Arg Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln
        355                 360                 365

Ala Leu Ser Gly Arg Leu Arg Ala Ile
    370                 375
```

<210> 36
<211> 384
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(384)
<223> Human APOBEC-3G polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(60)
<223> Cytoplasmic localization signal

<220>
<221> MISC_FEATURE

<222> (65)..(100)
<223> Nucleic acid editing domain

<220>
<221> MISC_FEATURE
<222> (257)..(291)
<223> Nucleic acid editing domain

<400> 36

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1   5   10   15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
            20   25   30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
        35   40   45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr
    50   55   60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65   70   75   80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
                85   90   95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro
            100   105   110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
        115   120   125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
    130   135   140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145   150   155   160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
                165   170   175

```
Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
            180                 185                 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
        195                 200                 205

Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
    210                 215                 220

Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225                 230                 235                 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
                245                 250                 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
            260                 265                 270

Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
        275                 280                 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
    290                 295                 300

Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305                 310                 315                 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
                325                 330                 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
            340                 345                 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
        355                 360                 365

Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
    370                 375                 380
```

<210> 37
<211> 382

<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(382)
<223> Human APOBEC-3B polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (66)..(100)
<223> Nucleic acid editing domain

<220>
<221> MISC_FEATURE
<222> (251)..(289)
<223> Nucleic acid editing domain

<400> 37

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln
50 55 60

Tyr His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80

Pro Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95

Cys Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ser Glu His Pro
100 105 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu
115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg

```
    130                 135                 140

Val Thr Ile Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145                 150                 155                 160

Val Tyr Asn Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu
                165                 170                 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu
            180                 185                 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val
        195                 200                 205

Leu Arg Arg Arg Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp
    210                 215                 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn
225                 230                 235                 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu
                245                 250                 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile
            260                 265                 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
        275                 280                 285

Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg
    290                 295                 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys
305                 310                 315                 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
                325                 330                 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
            340                 345                 350
```

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
        355   360   365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
    370   375   380

<210> 38
<211> 395
<212> PRT
<213> Rattus

<220>
<221> MISC_FEATURE
<222> (1)..(395)
<223> Rat APOBEC-3B polypeptide sequence

<400> 38

Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys
1   5   10   15

Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu
            20   25   30

Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg
        35   40   45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly
    50   55   60

Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys
65   70   75   80

Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp
                85   90   95

Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp
            100   105   110

Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg
        115   120   125

```
Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg
    130                 135                 140

Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
145                 150                 155                 160

Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
                165                 170                 175

Lys Lys Cys Trp Asn Lys Phe Val Asp Asn Asp Gly Gln Pro Phe Arg
            180                 185                 190

Pro Trp Met Arg Leu Arg Ile Asn Phe Ser Phe Tyr Asp Cys Lys Leu
        195                 200                 205

Gln Glu Ile Phe Ser Arg Met Asn Leu Leu Arg Glu Asp Val Phe Tyr
    210                 215                 220

Leu Gln Phe Asn Asn Ser His Arg Val Lys Pro Val Gln Asn Arg Tyr
225                 230                 235                 240

Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly
                245                 250                 255

Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His
            260                 265                 270

Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln
        275                 280                 285

Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
    290                 295                 300

Arg Gln Leu Ala Ala Phe Lys Lys Asp His Pro Asp Leu Ile Leu Arg
305                 310                 315                 320

Ile Tyr Thr Ser Arg Leu Tyr Phe Trp Arg Lys Lys Phe Gln Lys Gly
                325                 330                 335

Leu Cys Thr Leu Trp Arg Ser Gly Ile His Val Asp Val Met Asp Leu
            340                 345                 350
```

Pro Gln Phe Ala Asp Cys Trp Thr Asn Phe Val Asn Pro Gln Arg Pro
355 360 365

Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln Arg
370 375 380

Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu
385 390 395

<210> 39
<211> 226
<212> PRT
<213> Bos Taurus

<220>
<221> MISC_FEATURE
<222> (1)..(226)
<223> Bovine APOBEC-3B polypeptide sequence

<400> 39

Asp Gly Trp Glu Val Ala Phe Arg Ser Gly Thr Val Leu Lys Ala Gly
1 5 10 15

Val Leu Gly Val Ser Met Thr Glu Gly Trp Ala Gly Ser Gly His Pro
20 25 30

Gly Gln Gly Ala Cys Val Trp Thr Pro Gly Thr Arg Asn Thr Met Asn
35 40 45

Leu Leu Arg Glu Val Leu Phe Lys Gln Gln Phe Gly Asn Gln Pro Arg
50 55 60

Val Pro Ala Pro Tyr Tyr Arg Arg Lys Thr Tyr Leu Cys Tyr Gln Leu
65 70 75 80

Lys Gln Arg Asn Asp Leu Thr Leu Asp Arg Gly Cys Phe Arg Asn Lys
85 90 95

Lys Gln Arg His Ala Glu Arg Phe Ile Asp Lys Ile Asn Ser Leu Asp
100 105 110

Leu Asn Pro Ser Gln Ser Tyr Lys Ile Ile Cys Tyr Ile Thr Trp Ser
115 120 125

Pro Cys Pro Asn Cys Ala Asn Glu Leu Val Asn Phe Ile Thr Arg Asn
130 135 140

Asn His Leu Lys Leu Glu Ile Phe Ala Ser Arg Leu Tyr Phe His Trp
145 150 155 160

Ile Lys Ser Phe Lys Met Gly Leu Gln Asp Leu Gln Asn Ala Gly Ile
165 170 175

Ser Val Ala Val Met Thr His Thr Glu Phe Glu Asp Cys Trp Glu Gln
180 185 190

Phe Val Asp Asn Gln Ser Arg Pro Phe Gln Pro Trp Asp Lys Leu Glu
195 200 205

Gln Tyr Ser Ala Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr Ala
210 215 220

Pro Ile
225

<210> 40
<211> 490
<212> PRT
<213> Pan troglodytes

<220>
<221> MISC_FEATURE
<222> (1)..(490)
<223> Chimpanzee APOBEC-3B polypeptide sequence

<400> 40

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Trp Met Tyr Gln Arg Thr
1 5 10 15

Phe Tyr Tyr Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30

```
Thr Trp Leu Cys Tyr Glu Val Lys Ile Arg Arg Gly His Ser Asn Leu
        35                  40                  45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Met Tyr Ser Gln Pro Glu
    50                  55                  60

His His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65                  70                  75                  80

Ser Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
                85                  90                  95

Cys Pro Asp Cys Val Ala Lys Leu Ala Lys Phe Leu Ala Glu His Pro
            100                 105                 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu
        115                 120                 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg
    130                 135                 140

Val Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145                 150                 155                 160

Val Tyr Asn Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp
                165                 170                 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Ile Arg His Leu
            180                 185                 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val
        195                 200                 205

Leu Arg Arg His Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp
    210                 215                 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn
225                 230                 235                 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu
                245                 250                 255
```

```
Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile
            260                 265                 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly
        275                 280                 285

Cys Ala Gly Gln Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg
    290                 295                 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys
305                 310                 315                 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met
                325                 330                 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln
            340                 345                 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
        355                 360                 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Val Arg Ala Ser Ser Leu
    370                 375                 380

Cys Met Val Pro His Arg Pro Pro Pro Pro Pro Gln Ser Pro Gly Pro
385                 390                 395                 400

Cys Leu Pro Leu Cys Ser Glu Pro Pro Leu Gly Ser Leu Leu Pro Thr
                405                 410                 415

Gly Arg Pro Ala Pro Ser Leu Pro Phe Leu Leu Thr Ala Ser Phe Ser
            420                 425                 430

Phe Pro Pro Pro Ala Ser Leu Pro Pro Leu Pro Ser Leu Ser Leu Ser
        435                 440                 445

Pro Gly His Leu Pro Val Pro Ser Phe His Ser Leu Thr Ser Cys Ser
    450                 455                 460
```

```
Ile Gln Pro Pro Cys Ser Ser Arg Ile Arg Glu Thr Glu Gly Trp Ala
465                 470                 475                 480


Ser Val Ser Lys Glu Gly Arg Asp Leu Gly
                485                 490


<210>  41
<211>  190
<212>  PRT
<213>  Homo sapiens


<220>
<221>  MISC_FEATURE
<222>  (1)..(190)
<223>  Human APOBEC-3C polypeptide sequence

<220>
<221>  MISC_FEATURE
<222>  (65)..(100)
<223>  Nucleic acid editing domain

<400>  41

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1               5                   10                  15


Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
            20                  25                  30


Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
        35                  40                  45


Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
    50                  55                  60


Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu
65                  70                  75                  80


Ser Pro Asn Thr Lys Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
                85                  90                  95


Cys Pro Asp Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
            100                 105                 110
```

```
Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Tyr
        115                 120                 125


Pro Cys Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
    130                 135                 140


Val Glu Ile Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145                 150                 155                 160


Val Tyr Asn Asp Asn Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr
                165                 170                 175


Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ser Leu Gln
            180                 185                 190


<210>  42
<211>  190
<212>  PRT
<213>  Gorilla


<220>
<221>  MISC_FEATURE
<222>  (1)..(190)
<223>  Gorilla APOBEC-3C polypeptide sequence

<220>
<221>  MISC_FEATURE
<222>  (65)..(100)
<223>  Nucleic acid editing domain

<400>  42

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1               5                   10                  15


Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
            20                  25                  30


Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
        35                  40                  45


Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
    50                  55                  60
```

```
Cys His Ala Glu Arg Cys Phe Leu Ser Trp Glu Cys Asp Asp Ile Leu
65                  70                  75                  80


Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
                85                  90                  95


Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
            100                 105                 110


Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Asp
        115                 120                 125


Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
    130                 135                 140


Val Lys Ile Met Asp Tyr Lys Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145                 150                 155                 160


Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
                165                 170                 175


Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
            180                 185                 190


<210>  43
<211>  202
<212>  PRT
<213>  Rhesus macaque


<220>
<221>  MISC_FEATURE
<222>  (1)..(202)
<223>  Rhesus macaque APOBEC-3A polypeptide sequence

<220>
<221>  MISC_FEATURE
<222>  (72)..(103)
<223>  Nucleic acid editing domain

<400>  43

Met Asp Gly Ser Pro Ala Ser Arg Pro Arg His Leu Met Asp Pro Asn
1               5                   10                  15
```

```
Thr Phe Thr Phe Asn Phe Asn Asn Asp Leu Ser Val Arg Gly Arg His
            20                  25                  30

Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Trp
        35                  40                  45

Val Pro Met Asp Glu Arg Arg Gly Phe Leu Cys Asn Lys Ala Lys Asn
    50                  55                  60

Val Pro Cys Gly Asp Tyr Gly Cys His Val Glu Leu Arg Phe Leu Cys
65                  70                  75                  80

Glu Val Pro Ser Trp Gln Leu Asp Pro Ala Gln Thr Tyr Arg Val Thr
                85                  90                  95

Trp Phe Ile Ser Trp Ser Pro Cys Phe Arg Arg Gly Cys Ala Gly Gln
            100                 105                 110

Val Arg Val Phe Leu Gln Glu Asn Lys His Val Arg Leu Arg Ile Phe
        115                 120                 125

Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Gln Glu Ala Leu Arg
    130                 135                 140

Thr Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Glu Glu
145                 150                 155                 160

Phe Lys His Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe
                165                 170                 175

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg
            180                 185                 190

Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
        195                 200
```

<210> 44
<211> 185
<212> PRT
<213> Bos Taurus

<220>
<221> MISC_FEATURE
<222> (1)..(185)
<223> Bovine APOBEC-3A polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (56)..(86)
<223> Nucleic acid editing domain

<400> 44

Met Asp Glu Tyr Thr Phe Thr Glu Asn Phe Asn Asn Gln Gly Trp Pro
1 5 10 15

Ser Lys Thr Tyr Leu Cys Tyr Glu Met Glu Arg Leu Asp Gly Asp Ala
20 25 30

Thr Ile Pro Leu Asp Glu Tyr Lys Gly Phe Val Arg Asn Lys Gly Leu
35 40 45

Asp Gln Pro Glu Lys Pro Cys His Ala Glu Leu Tyr Phe Leu Gly Lys
50 55 60

Ile His Ser Trp Asn Leu Asp Arg Asn Gln His Tyr Arg Leu Thr Cys
65 70 75 80

Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Gln Lys Leu Thr Thr
85 90 95

Phe Leu Lys Glu Asn His His Ile Ser Leu His Ile Leu Ala Ser Arg
100 105 110

Ile Tyr Thr His Asn Arg Phe Gly Cys His Gln Ser Gly Leu Cys Glu
115 120 125

Leu Gln Ala Ala Gly Ala Arg Ile Thr Ile Met Thr Phe Glu Asp Phe
130 135 140

Lys His Cys Trp Glu Thr Phe Val Asp His Lys Gly Lys Pro Phe Gln
145 150 155 160

Pro Trp Glu Gly Leu Asn Val Lys Ser Gln Ala Leu Cys Thr Glu Leu
165 170 175

Gln Ala Ile Leu Lys Thr Gln Gln Asn
180 185

<210> 45
<211> 200
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(200)
<223> Human APOBEC-3H polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (54)..(88)
<223> Nucleic acid editing domain

<400> 45

Met Ala Leu Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys
1 5 10 15

Arg Arg Leu Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr
20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu
35 40 45

Asn Lys Lys Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys
50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu
65 70 75 80

Thr Trp Ser Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile
85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr
100 105 110

Tyr His Trp Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly
115 120 125

```
Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Lys Phe Ala Asp Cys
    130                 135                 140

Trp Glu Asn Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr
145                 150                 155                 160

Lys Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg
                165                 170                 175

Leu Glu Arg Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met
            180                 185                 190

Asp Ile Leu Cys Asp Ala Glu Val
        195                 200
```

<210> 46
<211> 210
<212> PRT
<213> Rhesus macaque

<220>
<221> MISC_FEATURE
<222> (1)..(210)
<223> Rhesus macaque APOBEC-3H polypeptide sequence

<400> 46

```
Met Ala Leu Leu Thr Ala Lys Thr Phe Ser Leu Gln Phe Asn Asn Lys
1               5                   10                  15

Arg Arg Val Asn Lys Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr
            20                  25                  30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly His Leu Lys
        35                  40                  45

Asn Lys Lys Lys Asp His Ala Glu Ile Arg Phe Ile Asn Lys Ile Lys
    50                  55                  60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu
65                  70                  75                  80
```

```
Thr Trp Ser Pro Cys Pro Ser Cys Ala Gly Glu Leu Val Asp Phe Ile
                85                  90                  95

Lys Ala His Arg His Leu Asn Leu Arg Ile Phe Ala Ser Arg Leu Tyr
            100                 105                 110

Tyr His Trp Arg Pro Asn Tyr Gln Glu Gly Leu Leu Leu Leu Cys Gly
        115                 120                 125

Ser Gln Val Pro Val Glu Val Met Gly Leu Pro Glu Phe Thr Asp Cys
    130                 135                 140

Trp Glu Asn Phe Val Asp His Lys Glu Pro Pro Ser Phe Asn Pro Ser
145                 150                 155                 160

Glu Lys Leu Glu Glu Leu Asp Lys Asn Ser Gln Ala Ile Lys Arg Arg
                165                 170                 175

Leu Glu Arg Ile Lys Ser Arg Ser Val Asp Val Leu Glu Asn Gly Leu
            180                 185                 190

Arg Ser Leu Gln Leu Gly Pro Val Thr Pro Ser Ser Ser Ile Arg Asn
        195                 200                 205

Ser Arg
    210
```

<210> 47
<211> 386
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(386)
<223> Human APOBEC-3D polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (78)..(109)
<223> Nucleic acid editing domain

<220>

<221> MISC_FEATURE
<222> (262)..(296)
<223> Nucleic acid editing domain

<400> 47

```
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1               5                  10                  15


Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
            20                  25                  30


Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
        35                  40                  45


Leu Trp Asp Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln
    50                  55                  60


Ser Asn His Arg Gln Glu Val Tyr Phe Arg Phe Glu Asn His Ala Glu
65                  70                  75                  80


Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Arg Leu Pro Ala Asn Arg
                85                  90                  95


Arg Phe Gln Ile Thr Trp Phe Val Ser Trp Asn Pro Cys Leu Pro Cys
            100                 105                 110


Val Val Lys Val Thr Lys Phe Leu Ala Glu His Pro Asn Val Thr Leu
        115                 120                 125


Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Arg Asp Arg Asp Trp Arg
    130                 135                 140


Trp Val Leu Leu Arg Leu His Lys Ala Gly Ala Arg Val Lys Ile Met
145                 150                 155                 160


Asp Tyr Glu Asp Phe Ala Tyr Cys Trp Glu Asn Phe Val Cys Asn Glu
                165                 170                 175


Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Ser
            180                 185                 190
```

```
Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met
        195                 200                 205

Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Leu Lys Ala Cys
    210                 215                 220

Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val Thr Lys His
225                 230                 235                 240

His Ser Ala Val Phe Arg Lys Arg Gly Val Phe Arg Asn Gln Val Asp
                245                 250                 255

Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys
            260                 265                 270

Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr
        275                 280                 285

Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu
    290                 295                 300

Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys
305                 310                 315                 320

Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln
                325                 330                 335

Glu Gly Ala Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys
            340                 345                 350

Trp Lys Asn Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys
        355                 360                 365

Gly Leu Gln Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile
    370                 375                 380

Leu Gln
385
```

<210> 48
<211> 236

<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(236)
<223> Human APOBEC-1 polypeptide sequence

<400> 48

```
Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
1               5                   10                  15


Arg Ile Glu Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu
            20                  25                  30


Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
        35                  40                  45


Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
    50                  55                  60


Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met
65                  70                  75                  80


Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
                85                  90                  95


Ser Gln Ala Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu
            100                 105                 110


Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg
        115                 120                 125


Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
    130                 135                 140


Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145                 150                 155                 160


Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met
                165                 170                 175
```

```
Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
            180                 185                 190


Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu
        195                 200                 205


His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
    210                 215                 220


Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg
225                 230                 235


<210>  49
<211>  229
<212>  PRT
<213>  Mus musculus


<220>
<221>  MISC_FEATURE
<222>  (1)..(229)
<223>  Mouse APOBEC-1 polypeptide sequence

<400>  49

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 Val Trp Arg His Thr Ser Gln Asn Thr Ser Asn His Val Glu Val
    50                  55                  60


Asn Phe Leu Glu Lys Phe Thr Thr Glu Arg Tyr Phe Arg 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 His Pro Tyr Val Thr Leu
            100                 105                 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Thr Asp Gln 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 Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro
145                 150                 155                 160

Pro Ser Asn Glu Ala Tyr Trp Pro Arg Tyr Pro His Leu Trp Val Lys
                165                 170                 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
            180                 185                 190

Leu Lys Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
        195                 200                 205

Thr Leu Gln Thr Cys His Tyr Gln Arg Ile Pro Pro His Leu Leu Trp
    210                 215                 220

Ala Thr Gly Leu Lys
225
```

<210> 50
<211> 224
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(224)
<223> Human APOBEC-2 polypeptide sequence

<400> 50

```
Met Ala Gln Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln
1               5                   10                  15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
```

```
            20                  25                  30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
        35                  40                  45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
    50                  55                  60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65                  70                  75                  80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
                85                  90                  95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
            100                 105                 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
        115                 120                 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
    130                 135                 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145                 150                 155                 160

Glu Ile Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
                165                 170                 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
            180                 185                 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile
        195                 200                 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
    210                 215                 220

<210>  51
<211>  224
<212>  PRT
```

<213> Mus musculus

<220>
<221> MISC_FEATURE
<222> (1)..(224)
<223> Mouse APOBEC-2 polypeptide sequence

<400> 51

```
Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln
1               5                  10                  15

Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
            20                  25                  30

Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro
        35                  40                  45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
    50                  55                  60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Val Gln Ser Lys Gly
65                  70                  75                  80

Gly Gln Ala Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly
                85                  90                  95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
            100                 105                 110

Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
        115                 120                 125

Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
    130                 135                 140

Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro
145                 150                 155                 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
                165                 170                 175
```

```
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
            180                 185                 190


Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
        195                 200                 205


Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
    210                 215                 220


<210>  52
<211>  224
<212>  PRT
<213>  Rattus


<220>
<221>  MISC_FEATURE
<222>  (1)..(224)
<223>  Rat APOBEC-2 polypeptide sequence

<400>  52

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln
1               5                   10                  15


Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
            20                  25                  30


Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro
        35                  40                  45


Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
    50                  55                  60


Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
65                  70                  75                  80


Gly Gln Val Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly
                85                  90                  95


Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
            100                 105                 110


Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
```

```
        115                 120                 125

Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
    130                 135                 140

Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro
145                 150                 155                 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
                165                 170                 175

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

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
        195                 200                 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
    210                 215                 220
```

<210> 53
<211> 208
<212> PRT
<213> Petromyzon marinus

<220>
<221> MISC_FEATURE
<222> (1)..(208)
<223> Petromyzon marinus CDA1 (pmCDA1) polypeptide sequence

<400> 53

```
Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr
1               5                   10                  15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
            20                  25                  30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
        35                  40                  45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
    50                  55                  60
```

```
Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg
65                  70                  75                  80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
                85                  90                  95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
            100                 105                 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
        115                 120                 125

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

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145                 150                 155                 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
                165                 170                 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
            180                 185                 190

Phe Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
        195                 200                 205


<210>  54
<211>  381
<212>  PRT
<213>  Homo sapiens


<220>
<221>  MISC_FEATURE
<222>  (1)..(381)
<223>  Human APOBEC3G D316R D317R polypeptide sequence

<400>  54

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1               5                   10                  15
```

```
Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
            20                  25                  30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
        35                  40                  45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr
    50                  55                  60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65                  70                  75                  80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
                85                  90                  95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro
            100                 105                 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
        115                 120                 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
    130                 135                 140

Pro Arg Ala Thr Met Lys Phe Asn Tyr Asp Glu Phe Gln His Cys Trp
145                 150                 155                 160

Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn
                165                 170                 175

Leu Pro Lys Tyr Tyr Ile Leu Leu His Phe Met Leu Gly Glu Ile Leu
            180                 185                 190

Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu
        195                 200                 205

Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu
    210                 215                 220

Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe
```

225 230 235 240

Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His
245 250 255

Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu
260 265 270

Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
275 280 285

Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Lys His Val Ser
290 295 300

Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln
305 310 315 320

Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr
325 330 335

Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly
340 345 350

Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu
355 360 365

Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 55
<211> 182
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(182)
<223> Human APOBEC3G chain A polypeptide sequence

<400> 55

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Trp
1 5 10 15

```
Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His Asn
            20                  25                  30

Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn Gln
        35                  40                  45

Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu Cys
    50                  55                  60

Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp Tyr
65                  70                  75                  80

Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln
                85                  90                  95

Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys Ile
            100                 105                 110

Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu
        115                 120                 125

Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr Tyr Ser Glu
    130                 135                 140

Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe
145                 150                 155                 160

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser Gly Arg
                165                 170                 175

Leu Arg Ala Ile Leu Gln
            180
```

<210> 56
<211> 184
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE

<222> (1)..(184)
<223> Human APOBEC3G chain A D120R D121R polypeptide sequence

<400> 56

```
Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val
1               5                   10                  15


Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His
            20                  25                  30


Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn
        35                  40                  45


Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu
    50                  55                  60


Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp
65                  70                  75                  80


Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala
                85                  90                  95


Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys
            100                 105                 110


Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly
        115                 120                 125


Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Met Thr Tyr
    130                 135                 140


Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys
145                 150                 155                 160


Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser
                165                 170                 175


Gly Arg Leu Arg Ala Ile Leu Gln
            180
```

<210> 57

<211> 367
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(367)
<223> hAPOBEC-4 polypeptide sequence

<400> 57

Met Glu Pro Ile Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val
1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro
20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe
35 40 45

Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Phe Pro Gln Thr Lys
50 55 60

His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln
65 70 75 80

Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser
85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110

Asp Ser Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asn
115 120 125

Glu Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Ile Thr
130 135 140

Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160

Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175

```
Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly
            180                 185                 190

Ile Trp His Ser Val Leu His Ser Phe Ile Ser Gly Val Ser Gly Ser
        195                 200                 205

His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Ala Asp Arg His
    210                 215                 220

Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Tyr Phe Thr
225                 230                 235                 240

Asp Val Leu Leu Gln Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Glu
                245                 250                 255

Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Phe Phe
            260                 265                 270

Gln Met Pro Ser Gly Gln Leu Gln Pro Asn Leu Pro Pro Asp Leu Arg
        275                 280                 285

Ala Pro Val Val Phe Val Leu Val Pro Leu Arg Asp Leu Pro Pro Met
    290                 295                 300

His Met Gly Gln Asn Pro Asn Lys Pro Arg Asn Ile Val Arg His Leu
305                 310                 315                 320

Asn Met Pro Gln Met Ser Phe Gln Glu Thr Lys Asp Leu Gly Arg Leu
                325                 330                 335

Pro Thr Gly Arg Ser Val Glu Ile Val Glu Ile Thr Glu Gln Phe Ala
            340                 345                 350

Ser Ser Lys Glu Ala Asp Glu Lys Lys Lys Lys Lys Gly Lys Lys
        355                 360                 365
```

<210> 58
<211> 388
<212> PRT
<213> Rattus norvegicus

<220>
<221> MISC_FEATURE
<222> (1)..(388)
<223> rAPOBEC-4 polypeptide sequence

<400> 58

```
Met Glu Pro Leu Tyr Glu Glu Tyr Leu Thr His Ser Gly Thr Ile Val
1               5                  10                  15


Lys Pro Tyr Tyr Trp Leu Ser Val Ser Leu Asn Cys Thr Asn Cys Pro
            20                  25                  30


Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Pro Tyr Thr Glu Phe
        35                  40                  45


His Gln Thr Phe Gly Phe Pro Trp Ser Thr Tyr Pro Gln Thr Lys His
    50                  55                  60


Leu Thr Phe Tyr Glu Leu Arg Ser Ser Ser Gly Asn Leu Ile Gln Lys
65                  70                  75                  80


Gly Leu Ala Ser Asn Cys Thr Gly Ser His Thr His Pro Glu Ser Met
                85                  90                  95


Leu Phe Glu Arg Asp Gly Tyr Leu Asp Ser Leu Ile Phe His Asp Ser
            100                 105                 110


Asn Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asp Glu
        115                 120                 125


Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Met Asn Tyr
    130                 135                 140


Pro Glu Val Thr Leu Ser Val Phe Phe Ser Gln Leu Tyr His Thr Glu
145                 150                 155                 160


Asn Gln Phe Pro Thr Ser Ala Trp Asn Arg Glu Ala Leu Arg Gly Leu
                165                 170                 175


Ala Ser Leu Trp Pro Gln Val Thr Leu Ser Ala Ile Ser Gly Gly Ile
```

```
            180                 185                 190

Trp Gln Ser Ile Leu Glu Thr Phe Val Ser Gly Ile Ser Glu Gly Leu
        195                 200                 205

Thr Ala Val Arg Pro Phe Thr Ala Gly Arg Thr Leu Thr Asp Arg Tyr
    210                 215                 220

Asn Ala Tyr Glu Ile Asn Cys Ile Thr Glu Val Lys Pro Tyr Phe Thr
225                 230                 235                 240

Asp Ala Leu His Ser Trp Gln Lys Glu Asn Gln Asp Gln Lys Val Trp
                245                 250                 255

Ala Ala Ser Glu Asn Gln Pro Leu His Asn Thr Thr Pro Ala Gln Trp
            260                 265                 270

Gln Pro Asp Met Ser Gln Asp Cys Arg Thr Pro Ala Val Phe Met Leu
        275                 280                 285

Val Pro Tyr Arg Asp Leu Pro Pro Ile His Val Asn Pro Ser Pro Gln
    290                 295                 300

Lys Pro Arg Thr Val Val Arg His Leu Asn Thr Leu Gln Leu Ser Ala
305                 310                 315                 320

Ser Lys Val Lys Ala Leu Arg Lys Ser Pro Ser Gly Arg Pro Val Lys
                325                 330                 335

Lys Glu Glu Ala Arg Lys Gly Ser Thr Arg Ser Gln Glu Ala Asn Glu
            340                 345                 350

Thr Asn Lys Ser Lys Trp Lys Lys Gln Thr Leu Phe Ile Lys Ser Asn
        355                 360                 365

Ile Cys His Leu Leu Glu Arg Glu Gln Lys Lys Ile Gly Ile Leu Ser
    370                 375                 380

Ser Trp Ser Val
385
```

<210> 59
<211> 363
<212> PRT
<213> Macaca fascicularis

<220>
<221> MISC_FEATURE
<222> (1)..(363)
<223> mfAPOBEC-4 polypeptide sequence

<400> 59

Met Glu Pro Thr Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val
1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro
20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe
35 40 45

Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Tyr Pro Gln Thr Lys
50 55 60

His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln
65 70 75 80

Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser
85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110

Asp Ser Ile Arg His Ile Ile Leu Tyr Cys Asn Asn Ser Pro Cys Asn
115 120 125

Glu Ala Asn His Cys Cys Ile Ser Lys Val Tyr Asn Phe Leu Ile Thr
130 135 140

Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160

```
Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
                165                 170                 175

Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly
            180                 185                 190

Ile Trp His Ser Val Leu His Ser Phe Val Ser Gly Val Ser Gly Ser
        195                 200                 205

His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Thr Asp Arg Tyr
    210                 215                 220

Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Phe Phe Thr
225                 230                 235                 240

Asp Val Leu Leu His Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Met
                245                 250                 255

Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Ser Phe
            260                 265                 270

Gln Met Thr Ser Gly Ile Pro Pro Asp Leu Arg Ala Pro Val Val Phe
        275                 280                 285

Val Leu Leu Pro Leu Arg Asp Leu Pro Pro Met His Met Gly Gln Asp
    290                 295                 300

Pro Asn Lys Pro Arg Asn Ile Ile Arg His Leu Asn Met Pro Gln Met
305                 310                 315                 320

Ser Phe Gln Glu Thr Lys Asp Leu Glu Arg Leu Pro Thr Arg Arg Ser
                325                 330                 335

Val Glu Thr Val Glu Ile Thr Glu Arg Phe Ala Ser Ser Lys Gln Ala
            340                 345                 350

Glu Glu Lys Thr Lys Lys Lys Lys Gly Lys Lys
        355                 360
```

<210> 60
<211> 218

<212> PRT
<213> Rattus norvegicus

<220>
<221> MISC_FEATURE
<222> (1)..(218)
<223> rAPOBEC-1 (delta 177-186) polypeptide sequence

<400> 60

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

```
Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu
            180                 185                 190


Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro
        195                 200                 205


Pro His Ile Leu Trp Ala Thr Gly Leu Lys
    210                 215


<210>  61
<211>  218
<212>  PRT
<213>  Rattus norvegicus


<220>
<221>  MISC_FEATURE
<222>  (1)..(218)
<223>  rAPOBEC-1 (delta 202-213) polypeptide sequence

<400>  61

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 His Tyr Gln Arg Leu Pro
        195                 200                 205


Pro His Ile Leu Trp Ala Thr Gly Leu Lys
    210                 215


<210>  62
<211>  198
<212>  PRT
<213>  Mus musculus


<220>
<221>  MISC_FEATURE
<222>  (1)..(198)
<223>  mouse cytidine deaminase polypeptide sequence

<220>
<221>  MISC_FEATURE
<222>  (1)..(30)
<223>  Nuclear localization sequence

<220>
<221>  MISC_FEATURE
<222>  (183)..(198)
<223>  Nuclear export signal

<400>  62

Met Asp Ser Leu Leu Met Lys Gln Lys Lys Phe Leu Tyr His Phe Lys
1               5                   10                  15
```

```
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
            20                  25                  30

Val Lys Arg Arg Asp Ser Ala Thr Ser Cys Ser Leu Asp Phe Gly His
        35                  40                  45

Leu Arg Asn Lys Ser 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 Glu
                85                  90                  95

Phe Leu Arg Trp 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 Gly Ile Met Thr Phe Lys Asp Tyr
    130                 135                 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Arg Thr Phe Lys
145                 150                 155                 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Thr 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 Met Leu Gly Phe
        195
```

<210> 63
<211> 224
<212> PRT
<213> Petromyzon marinus

<220>
<221> MISC_FEATURE
<222> (1)..(224)
<223> pmCDA-1 polypeptide sequence

<400> 63

Met Ala Gly Tyr Glu Cys Val Arg Val Ser Glu Lys Leu Asp Phe Asp
1   5   10   15

Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His
            20   25   30

Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg
        35   40   45

Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His
    50   55   60

Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65   70   75   80

Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
                85   90   95

Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
            100   105   110

Gln Gly His Thr Leu Thr Met His Phe Ser Arg Ile Tyr Asp Arg Asp
        115   120   125

Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn
    130   135   140

Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145   150   155   160

Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
                165   170   175

Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu
            180   185   190

Val Arg Cys Ala Gly Met Arg Glu Ser Gly Ile Pro Leu His Leu Phe
        195   200   205

Thr Leu Gln Thr Pro Leu Leu Ser Gly Arg Val Val Trp Trp Arg Val
    210   215   220

<210> 64
<211> 331
<212> PRT
<213> Petromyzon marinus

<220>
<221> MISC_FEATURE
<222> (1)..(331)
<223> pmCDA-2 polypeptide sequence

<400> 64

Met Glu Leu Arg Glu Val Val Asp Cys Ala Leu Ala Ser Cys Val Arg
1   5   10   15

His Glu Pro Leu Ser Arg Val Ala Phe Leu Arg Cys Phe Ala Ala Pro
            20   25   30

Ser Gln Lys Pro Arg Gly Thr Val Ile Leu Phe Tyr Val Glu Gly Ala
        35   40   45

Gly Arg Gly Val Thr Gly Gly His Ala Val Asn Tyr Asn Lys Gln Gly
    50   55   60

Thr Ser Ile His Ala Glu Val Leu Leu Leu Ser Ala Val Arg Ala Ala
65   70   75   80

Leu Leu Arg Arg Arg Arg Cys Glu Asp Gly Glu Glu Ala Thr Arg Gly
                85   90   95

Cys Thr Leu His Cys Tyr Ser Thr Tyr Ser Pro Cys Arg Asp Cys Val
            100   105   110

Glu Tyr Ile Gln Glu Phe Gly Ala Ser Thr Gly Val Arg Val Val Ile
        115   120   125

```
His Cys Cys Arg Leu Tyr Glu Leu Asp Val Asn Arg Arg Arg Ser Glu
    130                 135                 140

Ala Glu Gly Val Leu Arg Ser Leu Ser Arg Leu Gly Arg Asp Phe Arg
145                 150                 155                 160

Leu Met Gly Pro Arg Asp Ala Ile Ala Leu Leu Leu Gly Gly Arg Leu
                165                 170                 175

Ala Asn Thr Ala Asp Gly Glu Ser Gly Ala Ser Gly Asn Ala Trp Val
            180                 185                 190

Thr Glu Thr Asn Val Val Glu Pro Leu Val Asp Met Thr Gly Phe Gly
        195                 200                 205

Asp Glu Asp Leu His Ala Gln Val Gln Arg Asn Lys Gln Ile Arg Glu
    210                 215                 220

Ala Tyr Ala Asn Tyr Ala Ser Ala Val Ser Leu Met Leu Gly Glu Leu
225                 230                 235                 240

His Val Asp Pro Asp Lys Phe Pro Phe Leu Ala Glu Phe Leu Ala Gln
                245                 250                 255

Thr Ser Val Glu Pro Ser Gly Thr Pro Arg Glu Thr Arg Gly Arg Pro
            260                 265                 270

Arg Gly Ala Ser Ser Arg Gly Pro Glu Ile Gly Arg Gln Arg Pro Ala
        275                 280                 285

Asp Phe Glu Arg Ala Leu Gly Ala Tyr Gly Leu Phe Leu His Pro Arg
    290                 295                 300

Ile Val Ser Arg Glu Ala Asp Arg Glu Glu Ile Lys Arg Asp Leu Ile
305                 310                 315                 320

Val Val Met Arg Lys His Asn Tyr Gln Gly Pro
                325                 330
```

<210> 65

<211> 209
<212> PRT
<213> Petromyzon marinus

<220>
<221> MISC_FEATURE
<222> (1)..(209)
<223> pmCDA-5 polypeptide sequence

<400> 65

Met Ala Gly Asp Glu Asn Val Arg Val Ser Glu Lys Leu Asp Phe Asp
1 5 10 15

Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His
20 25 30

Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg
35 40 45

Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His
50 55 60

Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65 70 75 80

Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
85 90 95

Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
100 105 110

Gln Gly His Thr Leu Met Met His Phe Ser Arg Ile Tyr Asp Arg Asp
115 120 125

Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn
130 135 140

Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145 150 155 160

Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
165 170 175

Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu
            180   185   190

Val Arg Cys Ala Gly Met Arg Glu Ser Gly Met Pro Leu His Leu Phe
        195   200   205

Thr

<210> 66
<211> 158
<212> PRT
<213> Saccharomyces cerevisiae

<220>
<221> MISC_FEATURE
<222> (1)..(158)
<223> yCD polypeptide sequence

<400> 66

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp
1   5   10   15

Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro
            20   25   30

Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg
        35   40   45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu
    50   55   60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys
65   70   75   80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly
                85   90   95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val
            100   105   110

```
Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu
        115                 120                 125


Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe
    130                 135                 140


Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
145                 150                 155


<210>  67
<211>  194
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: cytidine deaminase
       polypeptide sequence

<400>  67

Met Lys Pro Gln Ile Arg Asp His Arg Pro Asn Pro Met Glu Ala Met
1               5                   10                  15


Tyr Pro His Ile Phe Tyr Phe His Phe Glu Asn Leu Glu Lys Ala Tyr
            20                  25                  30


Gly Arg Asn Glu Thr Trp Leu Cys Phe Thr Val Glu Ile Ile Lys Gln
        35                  40                  45


Tyr Leu Pro Val Pro Trp Lys Lys Gly Val Phe Arg Asn Gln Val Asp
    50                  55                  60


Pro Glu Thr His Cys His Ala Glu Lys Cys Phe Leu Ser Trp Phe Cys
65                  70                  75                  80


Asn Asn Thr Leu Ser Pro Lys Lys Asn Tyr Gln Val Thr Trp Tyr Thr
                85                  90                  95


Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu
            100                 105                 110


Ala Glu His Ser Asn Val Lys Leu Thr Ile Tyr Thr Ala Arg Leu Tyr
        115                 120                 125
```

```
Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Glu
    130                 135                 140

Glu Gly Ala Ser Val Glu Ile Met Asp Tyr Glu Asp Phe Gln Tyr Cys
145                 150                 155                 160

Trp Glu Asn Phe Val Tyr Asp Asp Gly Glu Pro Phe Lys Arg Trp Lys
                165                 170                 175

Gly Leu Lys Tyr Asn Phe Gln Ser Leu Thr Arg Arg Leu Arg Glu Ile
            180                 185                 190

Leu Gln
```

<210> 68
<211> 192
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: cytidine deaminase polypeptide sequence

<400> 68

```
Met Ala Asp Ser Ser Glu Lys Met Arg Gly Gln Tyr Ile Ser Arg Asp
1               5                   10                  15

Thr Phe Glu Lys Asn Tyr Lys Pro Ile Asp Gly Thr Lys Glu Ala His
            20                  25                  30

Leu Leu Cys Glu Ile Lys Trp Gly Lys Tyr Gly Lys Pro Trp Leu His
        35                  40                  45

Trp Cys Gln Asn Gln Arg Met Asn Ile His Ala Glu Asp Tyr Phe Met
    50                  55                  60

Asn Asn Ile Phe Lys Ala Lys Lys His Pro Val His Cys Tyr Val Thr
65                  70                  75                  80

Trp Tyr Leu Ser Trp Ser Pro Cys Ala Asp Cys Ala Ser Lys Ile Val
```

```
                85                  90                  95

Lys Phe Leu Glu Glu Arg Pro Tyr Leu Lys Leu Thr Ile Tyr Val Ala
            100                 105                 110

Gln Leu Tyr Tyr His Thr Glu Glu Glu Asn Arg Lys Gly Leu Arg Leu
        115                 120                 125

Leu Arg Ser Lys Lys Val Ile Ile Arg Val Met Asp Ile Ser Asp Tyr
    130                 135                 140

Asn Tyr Cys Trp Lys Val Phe Val Ser Asn Gln Asn Gly Asn Glu Asp
145                 150                 155                 160

Tyr Trp Pro Leu Gln Phe Asp Pro Trp Val Lys Glu Asn Tyr Ser Arg
                165                 170                 175

Leu Leu Asp Ile Phe Trp Glu Ser Lys Cys Arg Ser Pro Asn Pro Trp
            180                 185                 190
```

<210> 69
<211> 280
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: cytidine deaminase polypeptide sequence

<400> 69

```
Met Asp Pro Gln Arg Leu Arg Gln Trp Pro Gly Pro Gly Pro Ala Ser
1               5                   10                  15

Arg Gly Gly Tyr Gly Gln Arg Pro Arg Ile Arg Asn Pro Glu Glu Trp
            20                  25                  30

Phe His Glu Leu Ser Pro Arg Thr Phe Ser Phe His Phe Arg Asn Leu
        35                  40                  45

Arg Phe Ala Ser Gly Arg Asn Arg Ser Tyr Ile Cys Cys Gln Val Glu
    50                  55                  60
```

```
Gly Lys Asn Cys Phe Phe Gln Gly Ile Phe Gln Asn Gln Val Pro Pro
65                  70                  75                  80

Asp Pro Pro Cys His Ala Glu Leu Cys Phe Leu Ser Trp Phe Gln Ser
                85                  90                  95

Trp Gly Leu Ser Pro Asp Glu His Tyr Tyr Val Thr Trp Phe Ile Ser
            100                 105                 110

Trp Ser Pro Cys Cys Glu Cys Ala Ala Lys Val Ala Gln Phe Leu Glu
        115                 120                 125

Glu Asn Arg Asn Val Ser Leu Ser Leu Ser Ala Ala Arg Leu Tyr Tyr
    130                 135                 140

Phe Trp Lys Ser Glu Ser Arg Glu Gly Leu Arg Arg Leu Ser Asp Leu
145                 150                 155                 160

Gly Ala Gln Val Gly Ile Met Ser Phe Gln Asp Phe Gln His Cys Trp
                165                 170                 175

Asn Asn Phe Val His Asn Leu Gly Met Pro Phe Gln Pro Trp Lys Lys
            180                 185                 190

Leu His Lys Asn Tyr Gln Arg Leu Val Thr Glu Leu Lys Gln Ile Leu
        195                 200                 205

Arg Glu Glu Pro Ala Thr Tyr Gly Ser Pro Gln Ala Gln Gly Lys Val
    210                 215                 220

Arg Ile Gly Ser Thr Ala Ala Gly Leu Arg His Ser His Ser His Thr
225                 230                 235                 240

Arg Ser Glu Ala His Leu Arg Pro Asn His Ser Ser Arg Gln His Arg
                245                 250                 255

Ile Leu Asn Pro Pro Arg Glu Ala Arg Ala Arg Thr Cys Val Leu Val
            260                 265                 270

Asp Ala Ser Trp Ile Cys Tyr Arg
        275                 280
```

<210> 70
<211> 555
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.1;
tr|F7B644|F7B644_HORSE polypeptide sequence

<400> 70

Met Pro Arg Gly Arg Ala Arg Glu Arg Gln Arg Arg Asn Pro Met Glu
1 5 10 15

Lys Leu Asp Ala Glu Ala Phe Ser Phe His Phe Leu Asn Met Glu Phe
20 25 30

Val Tyr Asp Arg Asn Cys Ser Tyr Leu Cys Tyr Gln Val Glu Gly Arg
35 40 45

Leu Ser Gly Ser Pro Val Leu Ser Glu Gln Gly Val Phe Pro Asn Glu
50 55 60

Val Cys Gly Lys Thr Arg Arg His Ala Glu Leu Cys Phe Leu Asp Trp
65 70 75 80

Phe Arg Gly Arg Leu Ser Pro Asp Glu Tyr Tyr Cys Val Thr Trp Phe
85 90 95

Ile Ser Trp Ser Pro Cys Ser Asn Cys Ala Arg Glu Val Ala Glu Phe
100 105 110

Leu Lys Arg His Arg Asn Val Glu Leu Ser Ile Phe Ala Ala Arg Leu
115 120 125

Tyr Tyr Cys Arg Asp His Glu Gln Gly Leu Gln Ser Leu Cys Asn Arg
130 135 140

Gly Ala Gln Leu Ala Val Met Leu Arg Lys Asp Phe Thr Tyr Cys Trp
145 150 155 160

Asp Asn Phe Val His Asn Ser Gly Arg Glu Phe Ser Pro Trp Glu Asn

```
                165                 170                 175

Ile Asp Ala Asn Ser Asp Leu Leu Ala Arg Lys Leu Glu Asp Leu Leu
            180                 185                 190

Lys Asn Pro Met Glu Lys Leu His Arg Lys Thr Phe Ser Phe His Phe
        195                 200                 205

Arg Asn Leu Lys Phe Ala Lys Gly Arg Lys Cys Ser Tyr Leu Cys Tyr
    210                 215                 220

Arg Val Glu Gly Arg Leu Ser Gly Ser Pro Gly Leu Ser Glu Gln Gly
225                 230                 235                 240

Val Phe Leu Asn Glu Val Cys Asp Glu Asn Cys Arg His Ala Glu Leu
                245                 250                 255

Cys Phe Leu His Trp Phe Arg Gly Arg Leu Ser Pro His Ala Asp Tyr
            260                 265                 270

Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Ser Asn Cys Ala Arg
        275                 280                 285

Glu Val Ala Glu Phe Leu Lys Gln His Arg Asn Val Glu Leu His Ile
    290                 295                 300

Ser Ala Ala Arg Leu Tyr Tyr Trp Gln Arg Asn Lys Pro Gly Leu Arg
305                 310                 315                 320

Asn Leu Arg Ser Ser Gly Ala Gln Leu Ala Ile Met Phe Phe Trp Asp
                325                 330                 335

Phe Arg Asp Cys Trp Asp Asn Phe Val His Asn Ser Gly Arg His Phe
            340                 345                 350

Ile Pro Trp Lys Lys Ile Asn Val Asn Ser Arg Leu Leu Ala Thr Lys
        355                 360                 365

Leu Glu Asp Leu Leu Lys Asn Pro Leu Glu Lys Leu His Pro Asn Thr
    370                 375                 380
```

```
Phe Ser Phe His Phe Cys Asn Leu Glu Phe Ala Tyr Asp Arg Lys Tyr
385                 390                 395                 400

Ser Tyr Leu Cys Tyr Gln Val Glu Gly Arg Leu Ser Gly Ser Pro Gly
                405                 410                 415

Leu Ser Glu Gln Gly Val Phe Leu Asn Glu Val Cys Gly Lys Thr Arg
            420                 425                 430

Cys His Ala Glu Leu Cys Phe Leu Asp Trp Phe Arg Val Arg Leu Ser
        435                 440                 445

Pro Asp Glu Tyr Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys
    450                 455                 460

Phe Tyr Cys Ala Arg Glu Val Ala Asp Phe Leu Lys Gln Tyr Arg Asn
465                 470                 475                 480

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

Ala Gln Gly Leu Arg Ser Leu Cys Ser Ser Gly Ala Gln Leu Ala Ile
            500                 505                 510

Met Phe Phe Trp Asp Phe Arg Tyr Cys Trp Asp Asn Phe Val His Asn
        515                 520                 525

Ser Gly Arg Glu Phe Arg Pro Trp Lys Lys Ile Asn Val Asn Ser Arg
    530                 535                 540

Leu Leu Ala Thr Lys Leu Glu Asp Ile Leu Lys
545                 550                 555
```

<210> 71
<211> 437
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.2; tr|D1LZA1|D1LZA1_PANTI polypeptide sequence

<400> 71

```
Met Glu Pro Trp Arg Pro Ser Pro Arg Asn Pro Met Asp Arg Ile Asp
1               5                  10                  15

Pro Lys Thr Phe Arg Phe Gln Phe Pro Asn Leu Arg Tyr Ala Ser Gly
            20                  25                  30

Arg Lys Leu Cys Tyr Leu Cys Phe Gln Val Glu Arg Asp Tyr Phe Tyr
        35                  40                  45

Tyr Asn Asp Ser Asp Trp Gly Val Phe Arg Asn Glu Val His Pro Trp
    50                  55                  60

Ala Pro Cys His Ala Glu Gln Cys Phe Leu Ser Trp Phe Arg Asp Gln
65                  70                  75                  80

Tyr Pro Tyr Arg Asp Glu Asp Tyr Asn Val Thr Trp Phe Leu Ser Trp
                85                  90                  95

Ser Pro Cys Pro Thr Cys Ala Glu Glu Val Val Glu Phe Leu Glu Glu
            100                 105                 110

Tyr Arg Asn Leu Thr Leu Ser Ile Phe Thr Ser Arg Leu Tyr Tyr Phe
        115                 120                 125

Trp His Pro Asn Tyr Gln Glu Gly Leu Cys Lys Leu Trp Asp Ala Gly
    130                 135                 140

Val Gln Leu Asp Ile Met Ser Cys Asp Glu Phe Glu Tyr Cys Trp Asp
145                 150                 155                 160

Asn Phe Val Tyr His Lys Gly Met Arg Phe Gln Arg Arg Asn Leu Leu
                165                 170                 175

Lys Asp Tyr Asp Phe Leu Ala Ala Lys Leu Gln Glu Ile Leu Ser Pro
            180                 185                 190

Gly Gln Gln Arg Lys Arg Asp Trp Pro Phe Pro Pro Arg Pro Gly Ala
        195                 200                 205
```

```
Gln Val Asp Pro Arg Ser Trp Val Gln Glu Val Thr Glu Pro Gly Ile
    210                 215                 220

Asn Thr Arg Arg His Pro Leu His Leu Leu Val Ser Phe Leu Leu Pro
225                 230                 235                 240

Arg Pro Thr Met Asn Pro Leu Gln Glu Asp Ile Phe Tyr Arg Gln Phe
                245                 250                 255

Gly Asn Gln His Arg Val Pro Lys Pro Tyr Tyr Tyr Arg Arg Lys Thr
            260                 265                 270

Tyr Leu Cys Tyr Gln Leu Lys Leu Pro Glu Gly Thr Leu Ile Asp Lys
        275                 280                 285

Asp Cys Leu Arg Asn Lys Lys Lys Arg His Ala Glu Ile Cys Phe Ile
    290                 295                 300

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

Ile Cys Tyr Ile Thr Trp Ser Pro Cys Pro Phe Cys Ala Glu Glu Leu
                325                 330                 335

Val Ala Phe Val Lys Asp Asn Pro His Leu Ser Leu Arg Ile Phe Ala
            340                 345                 350

Ser Arg Leu Tyr Val His Trp Arg Trp Lys Tyr Gln Gln Gly Leu Arg
        355                 360                 365

His Leu His Ala Ser Gly Ile Pro Val Ala Val Met Ser Leu Pro Glu
    370                 375                 380

Phe Glu Asp Cys Trp Arg Asn Phe Val Asp His Gln Asp Arg Leu Phe
385                 390                 395                 400

Gln Pro Trp Arg Asn Leu Asp Gln Tyr Ser Glu Ser Ile Lys Arg Arg
                405                 410                 415

Leu Gly Lys Ile Leu Thr Pro Leu Asn Asp Leu Arg Asn Asp Phe Arg
            420                 425                 430
```

Asn Leu Lys Leu Glu
435

<210> 72
<211> 440
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.3;
tr|A0A3Q0DM17|A0A3Q0DM17_TARSY polypeptide sequence

<400> 72

Met Pro Met Lys Arg Met Tyr Ser Asn Ile Tyr Phe Asp His Phe Asn
1 5 10 15

Asn Gln Arg Leu Leu Ser Gly Gln Asn Ala Pro Trp Leu Cys Phe Lys
20 25 30

Val Glu Arg Val Glu Asn Cys Met Leu Val Pro Leu Glu Thr Gly Val
35 40 45

Phe Gly Asn Gln Val Ser Gly Cys Cys Gly Lys Thr Glu Arg Pro Val
50 55 60

Glu Pro Thr Ser Leu Thr Arg Ser Val Leu Val Ser Pro Asn Pro Gly
65 70 75 80

Thr Glu Leu Arg Ala Gln Gln Pro Ser Arg Lys Gly His Leu Gly Lys
85 90 95

Leu Gly Cys Val Glu Tyr Pro Ser Pro Gly Leu Ala Leu Val Met Leu
100 105 110

Gly Tyr Gly Ala Ser Thr Tyr Cys Pro Asp Ser Ser Met Tyr Cys Pro
115 120 125

Glu Thr Cys His His Pro Glu Met Cys Phe Leu Tyr Trp Phe Glu Lys
130 135 140

Thr Leu Ser His Glu Glu Gln Tyr Gln Ile Thr Trp Tyr Val Ser Trp

```
145                 150                 155                 160

Ser Pro Cys Val Asn Cys Ala Glu Glu Val Ala Glu Phe Leu Ser Val
                165                 170                 175

His Pro Lys Val Asn Leu Thr Ile Tyr Ala Ala Arg Leu Tyr Cys Tyr
            180                 185                 190

Gln Lys Leu Asn His Arg Gln Gly Leu Arg Arg Leu Cys Lys Glu Gly
        195                 200                 205

Ala Cys Val Lys Ile Met Asn Tyr Glu Glu Phe Asp His Cys Trp Glu
    210                 215                 220

Asn Phe Val Tyr Asn Asn Tyr Lys Ser Phe Lys Pro Trp Val Lys Leu
225                 230                 235                 240

Gln Asp Asn Tyr Glu Leu Leu Ala Thr Glu Leu Asp Lys Ile Leu Arg
                245                 250                 255

Ile Pro Met Glu Arg Met Pro Gln Lys Lys Phe Arg Phe His Phe Gln
            260                 265                 270

Asn Leu Ile Ala Lys Asp Arg Asn Thr Thr Trp Leu Cys Phe Glu Val
        275                 280                 285

Lys Asn Val Arg Lys Lys His Pro Pro Asp Leu Leu Glu Arg Gly Ile
    290                 295                 300

Phe Gln Asn Gln Val Thr Pro Arg Ile Asn Cys His Ala Glu Met Cys
305                 310                 315                 320

Phe Leu Ser Trp Phe Leu Glu Asn Met Leu Leu His Gly Lys Arg Tyr
                325                 330                 335

Gln Val Thr Trp Tyr Ile Ser Trp Ser Pro Cys Ser Ile Cys Ala Glu
            340                 345                 350

Glu Val Ala Glu Phe Leu Ser Ala His Pro Lys Val Ser Leu Thr Ile
        355                 360                 365
```

```
Tyr Ala Ala Arg Leu Tyr Tyr Phe Trp Val Pro Gly Tyr Arg Gln Gly
    370                 375                 380

Leu Arg Arg Leu Val Glu Glu Gly Ala Arg Val Glu Ile Met Asn Tyr
385                 390                 395                 400

Glu Glu Phe Asp Tyr Cys Trp Glu Asn Phe Val Ser Ile Asn Asn Glu
                405                 410                 415

Pro Phe Gln Pro Trp Glu Gly Leu His Glu Lys Tyr Gly Tyr Leu Val
            420                 425                 430

Thr Lys Leu Asn Asn Ile Leu Gly
        435                 440
```

<210> 73
<211> 453
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.4; tr|A0A3Q0DNJ5|A0A3Q0DNJ5_TARSY polypeptide sequence

<400> 73

```
Met Glu Asp Asn Pro Glu Pro Arg Pro Arg Gln Gln Met Asp Gln Asp
1               5                   10                  15

Thr Phe Ile Phe Asn Phe Asn Asn Asp Pro Ser Val Arg Gly Arg His
            20                  25                  30

Gln Thr Phe Leu Cys Tyr Glu Val Glu His Leu Asp Asp Asp Thr Trp
        35                  40                  45

Val Pro Gln Asp Lys Tyr Leu Gly Phe Leu His Asn Gln Pro Gln Ser
    50                  55                  60

Arg Ser Asn Ala Tyr Cys Ala Tyr His Ala Glu Leu Cys Phe Leu Glu
65                  70                  75                  80

Leu Val Ser Ser Trp Gln Leu Asp Pro Ala Gln Arg Tyr Arg Val Thr
                85                  90                  95
```

```
Cys Phe Ile Ser Trp Ser Pro Cys Ser Ser Cys Ala Gln Glu Val Ala
            100                 105                 110

Ala Phe Leu Lys Lys Asn Arg His Val Thr Leu Arg Ile Leu Ala Ala
        115                 120                 125

Arg Ile Tyr Asp Tyr Tyr Gln Gly Tyr Glu Asp Gly Leu Arg Thr Leu
    130                 135                 140

Gln Gly Val Gly Val Asp Ile Thr Val Met Thr Ser Ala Glu Phe Gly
145                 150                 155                 160

His Cys Trp Asn Thr Phe Val Asp His Gln Gly Ser Pro Phe Gln Pro
                165                 170                 175

Trp Glu Gly Leu Asp Gln His Ser Gln Val Ile Trp Gln Arg Met Gln
            180                 185                 190

Asp Ile Leu Gln Val Ile Pro Ala Lys Tyr Leu Met Glu Lys Val Lys
        195                 200                 205

Tyr Thr Val Thr Val Asp Ile Leu Phe Lys Gly Arg Val Pro Gly Pro
    210                 215                 220

Arg Tyr Leu Met Asp Gln Asn Thr Phe Thr Arg Asn Phe Ile Asn Asn
225                 230                 235                 240

Leu Ser Val Ser Gly Arg Arg Gln Thr Leu Leu Cys Tyr Glu Val Glu
                245                 250                 255

Arg Leu Gly Gly Asp Ile Trp Val Pro Leu Asp Gln Leu Arg Gly Phe
            260                 265                 270

Leu Leu Ser Gln Ala Arg Asp Val Leu Asn Tyr Tyr Gln Gly Arg His
        275                 280                 285

Ala Glu Pro Cys Phe Leu Asp Leu Val Ser Ser Trp Gln Leu Asp Pro
    290                 295                 300
```

```
Ala Gln His Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Thr
305                 310                 315                 320

Ser Cys Ala Gln Ala Val Ala Ala Phe Leu Arg Glu Asn Arg His Val
                325                 330                 335

Thr Leu Arg Ile Leu Ala Ala Arg Ile Tyr Asp Tyr His Gln Gly Tyr
            340                 345                 350

Glu Glu Gly Leu Arg Thr Leu Gln Arg Thr Gly Ala His Ile Asp Ile
        355                 360                 365

Met Thr Phe Lys Glu Phe Gly His Cys Trp Asn Thr Phe Val Asn His
    370                 375                 380

Lys Gly Ser Pro Phe Lys Ser Trp Thr Gly Leu Asp Gln His Ser Gln
385                 390                 395                 400

Ala Leu Arg Lys Arg Leu Gln Asp Ile Leu His Thr Met Ala Ser Ser
                405                 410                 415

Leu Trp Asp Gln Ser Glu Pro Lys Lys Pro Ile Pro Ser Gln Glu Val
            420                 425                 430

Thr Leu Pro Glu Ser Ile Pro Pro Ser His Gly Asn Arg Phe Arg Leu
        435                 440                 445

Val Lys Arg Pro Ser
    450
```

<210> 74
<211> 424
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.5;
       tr|G5AYU5|G5AYU5_HETGA polypeptide sequence

<400> 74

```
Phe Cys Phe Leu Ser Cys Val His Arg Lys Pro Ile Glu Arg Ile Tyr
1               5                   10                  15
```

```
Lys Lys Ala Phe Arg Phe Tyr Phe Arg Asn Leu Arg Cys Ala Tyr Gly
            20                  25                  30

Arg Asn Lys Thr Phe Leu Cys Tyr Glu Val Lys Arg Glu Arg Asp Asn
        35                  40                  45

Lys Val Leu His Lys Gly Val Val Leu Asn Gln Val Glu Pro Tyr Met
    50                  55                  60

Pro Leu His Ala Glu Leu Arg Phe Leu Ser Trp Phe His Asp Thr Leu
65                  70                  75                  80

Leu Cys Pro Leu Gly Ser Tyr Gln Val Thr Leu Tyr Val Ser Trp Ser
                85                  90                  95

Pro Cys Ser Glu Cys Ala Glu Glu Leu Thr Thr Phe Leu Ala Gly His
            100                 105                 110

Arg Asn Val Thr Met Thr Ile Tyr Val Ala Gln Leu Tyr Tyr Cys Asn
        115                 120                 125

Trp Lys Ser Pro Asn Arg Glu Gly Leu Lys Ile Leu Ile Ala Glu Asp
    130                 135                 140

Ala Arg Leu Arg Val Met Phe Tyr Asp Glu Phe Leu Tyr Cys Trp Arg
145                 150                 155                 160

Asn Phe Val Lys Asn Asp Tyr Asn Asn Phe Asp Pro Trp Ser Leu Leu
                165                 170                 175

Asp Glu Asn Ser Arg Tyr His Asn Arg Ile Leu Gln Asn Ile Leu Lys
            180                 185                 190

Gly Trp Gly Arg Pro His Arg Val Gly Pro Glu Gly Glu Gln Thr Ala
        195                 200                 205

Thr Pro Gly Gly Ser Gly Gly His Cys Ile Ser Val Phe Ser Leu Leu
    210                 215                 220

Arg Arg Arg Glu Met Thr Leu Lys Glu Glu Thr Phe Arg Val Gln Phe
```

```
225                 230                 235                 240

Asn Asn Ala Tyr Lys Ala Pro Lys Pro Tyr Arg Arg Arg Val Thr Tyr
                245                 250                 255

Leu Cys Tyr Gln Leu Gln Glu Ala Asn Gly Asp Pro Leu Thr Lys Gly
            260                 265                 270

Cys Leu Arg Thr Lys Lys Gly Tyr His Ala Glu Ser Arg Phe Ile Lys
        275                 280                 285

Arg Ile Cys Ser Met Asp Leu Gly Gln Asp Gln Ser Tyr Gln Val Thr
    290                 295                 300

Cys Phe Leu Thr Trp Ser Pro Cys Pro His Cys Ala Gln Glu Leu Val
305                 310                 315                 320

Ser Phe Lys Arg Ala His Pro His Leu Arg Leu Gln Ile Phe Thr Ala
                325                 330                 335

Arg Leu Phe Phe His Trp Lys Arg Ser Tyr Gln Glu Gly Leu Gln Arg
            340                 345                 350

Leu Cys Arg Ala Gln Val Pro Val Ala Val Met Gly His Pro Glu Phe
        355                 360                 365

Ala Tyr Cys Trp Asp Asn Phe Val Asp His Gln Pro Gly Pro Phe Glu
    370                 375                 380

Pro Pro Trp Ala Lys Leu Glu Tyr Tyr Ser Ser Cys Leu Lys Arg Arg
385                 390                 395                 400

Leu Gln Gln Ile Leu Arg Ser Trp Gly Val Asp Asp Leu Thr Asn Asp
                405                 410                 415

Phe Arg Asn Leu Gln Leu Gly Pro
            420


<210>  75
<211>  385
<212>  PRT
```

<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.6;
tr|A0A2Y9QMV5|A0A2Y9QMV5_TRIMA polypeptide sequence

<400> 75

Met Leu Ser Ser Pro Gln Thr Pro Gly Thr Arg Lys Pro Met Lys Thr
1 5 10 15

Leu Ala Pro Asp Glu Phe Ser Phe Asn Phe Glu Asn Leu Arg Leu Ala
20 25 30

His Gly Arg Asn Thr Thr Phe Leu Cys Phe Gln Val Glu Thr Lys Ala
35 40 45

Pro Pro Ser Leu Asn Ser Pro Asp Ser Gly Ile Phe Gln Asn Gln Asp
50 55 60

His Cys Pro Ser His His His Ala Glu Met Val Phe Leu Thr Trp Phe
65 70 75 80

Gln Lys Arg Leu Ser Pro Ala Gln His Tyr Glu Val Thr Trp Tyr Met
85 90 95

Ser Trp Ser Pro Cys Ser Arg Cys Ala Val Gln Val Ala Lys Phe Leu
100 105 110

Lys Ser Asn Ser Thr Val Asn Leu Ser Ile Phe Val Ala Arg Leu Tyr
115 120 125

Tyr Pro Arg Glu Leu Glu Thr Lys Asp Gly Leu His Ser Leu Trp Gln
130 135 140

Ala Gly Ala Gln Val Gln Ile Met Phe Phe Gln Asp Phe Lys Tyr Cys
145 150 155 160

Trp Glu Asn Phe Val Asn Asn Glu Gly Lys Pro Phe Gln Pro Trp Lys
165 170 175

Asn Leu Asp Glu Asn Ser Lys Asp Trp Asp Thr Glu Leu Lys Asp Ile
180 185 190

```
His Arg Asn Thr Thr Asp Leu Leu Thr Glu Glu Met Phe Tyr Ser Gln
        195                 200                 205

Phe Tyr Asn Arg Glu Lys Lys Ser Ser Ile Pro Arg Lys Thr Tyr Leu
    210                 215                 220

Cys Tyr Gln Leu Asn Glu Pro Gln Pro Val Lys Arg Cys Leu His Tyr
225                 230                 235                 240

Lys Lys Gly Tyr His Ala Val Thr Arg Phe Ile Asp Gly Ile Val Ser
                245                 250                 255

Met Asn Leu Asp Pro Ala Arg Ser Tyr Asp Ile Thr Cys Tyr Phe Thr
            260                 265                 270

Trp Ser Pro Cys Asn Arg Tyr Ala Arg Lys Leu Val Ser Phe Ile Glu
        275                 280                 285

Asp Tyr Pro Asn Leu Arg Leu Lys Val Tyr Thr Ser Arg Leu Tyr Phe
    290                 295                 300

His Trp Cys Trp Thr Asn Met Gln Gly Leu Gln His Leu Gln Asn Ser
305                 310                 315                 320

Arg Val Thr Val Ala Val Met Thr Phe Arg Asp Phe Glu Tyr Cys Trp
                325                 330                 335

Lys Asn Phe Val Asp Asn Gln Gly Lys Pro Phe Glu Pro Trp Glu Lys
            340                 345                 350

Leu Asp Leu Tyr Ser Gln Ser Thr Glu Arg Arg Leu Arg Arg Ile Leu
        355                 360                 365

Lys Pro Leu Thr Pro Asp Val Leu Asn Glu Asp Phe Gly Asn Leu His
    370                 375                 380

Leu
385
```

<210> 76
<211> 381
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.7;
tr|H0XHI0|H0XHI0_OTOGA polypeptide sequence

<400> 76

Leu Ser Cys Ala Phe Arg Asp Pro Met Asn Arg Met Tyr Pro Lys Thr
1 5 10 15

Phe Cys Gln Asn Phe Glu Lys Glu Pro Cys Pro Ser Asn Gln Asn Ser
20 25 30

Ser Trp Leu Cys Phe Glu Val Glu Thr Lys Asn Ser Ala Val Phe Phe
35 40 45

His Arg Gly Val Phe Arg Asn Gln Pro Ala Pro Pro Pro Arg Ala Pro
50 55 60

Thr Ser Val Leu Leu Ser Gln Gly Pro Val Lys Thr Pro Cys His Ala
65 70 75 80

Glu Glu Cys Phe Leu Thr Trp Ile Gln Gly Val Leu Pro Pro Asp His
85 90 95

His Tyr His Val Thr Trp Tyr Val Ser Arg Gly Pro Cys Ala Asn Cys
100 105 110

Ala Asn Leu Ile Val His Phe Leu Ala Met His Arg Arg Val Thr Leu
115 120 125

Thr Ile Phe Ala Ala His Leu Asn Phe Phe Trp Glu Ser Asp Phe Gln
130 135 140

Gln Gly Leu Leu Arg Met Asp Gln Glu Gly Val Gln Leu His Ile Met
145 150 155 160

Gly Tyr Glu Glu Phe Glu Tyr Cys Trp Asp Asn Phe Val Tyr Asn Gln
165 170 175

```
Arg Lys Gln Phe Val Pro Trp Asn Gly Leu Asn Glu Asn Tyr Glu Phe
            180                 185                 190

Met Val Ser Thr Leu Glu Asp Ile Leu Arg Ser Pro Leu Asp Arg Ile
        195                 200                 205

Arg Gln Lys Asp Phe Ser Ile His Phe Arg Asn Ser Leu Trp Leu Asp
    210                 215                 220

Asp Lys Ser Thr Trp Leu Cys Phe Glu Val Lys Arg Thr Lys Ser Pro
225                 230                 235                 240

Val Pro Leu Tyr Arg Gly Val Phe Arg Asn Gln Ser Pro Pro Lys Thr
                245                 250                 255

Pro Cys His Ala Glu Val Arg Phe Phe Thr Trp Leu Gln Asp Leu Pro
            260                 265                 270

Pro Asp Phe Cys Cys Gln Phe Thr Trp Tyr Leu Ser Trp Ser Pro Cys
        275                 280                 285

Ala Asp Cys Ala Asp Leu Val Ala Asn Phe Leu Ala Lys His Arg Asn
    290                 295                 300

Val Ser Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Tyr Arg Asp Pro
305                 310                 315                 320

Glu Met His Arg Gly Leu Arg Arg Met Tyr Gln Glu Gly Ala Asn Val
                325                 330                 335

Asp Ile Met Ser Val Ile Glu Phe Glu Tyr Cys Trp Asp Asn Phe Val
            340                 345                 350

Tyr Asn Gln Gly Lys Gln Phe Val Pro Trp Asn Gly Leu Asn Glu Asn
        355                 360                 365

Tyr Glu Phe Leu Val Pro Arg Leu Gln Glu Ile Leu Glu
    370                 375                 380
```

<210> 77

<211> 337
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.8;
tr|A0A3M0K4Y7|A0A3M0K4Y7_HIRRU polypeptide sequence

<400> 77

Met Tyr Ile Ser Lys Lys Ala Leu Arg Arg His Phe Asp Pro Arg Val
1 5 10 15

Tyr Pro Arg Glu Thr Tyr Leu Leu Cys Glu Leu Gln Trp Glu Gly Ser
20 25 30

Arg Arg Val Trp Ile His Trp Ile Arg Asn Val Pro Asp His His Ala
35 40 45

Glu Glu Tyr Phe Leu Glu Glu Val Phe Glu Pro Arg Asn Tyr Gly Phe
50 55 60

Cys Asn Ile Thr Leu Tyr Leu Ser Trp Ser Pro Cys Cys Thr Cys Cys
65 70 75 80

Ser Lys Ile Arg Asp Phe Leu Lys Arg Asn Pro Asn Val Lys Ile Asp
85 90 95

Ile Arg Val Ala Arg Leu Ile Tyr Pro Asp Tyr Ala Glu Thr Arg Ser
100 105 110

Ser Leu Arg Glu Leu Asn Gly Leu Gln Arg Val Ser Ile Gln Val Met
115 120 125

Glu Ala Ala Gly Leu Ser Cys Ile Glu Ser Lys Asn His Arg Ile Ser
130 135 140

Gln Val Glu Arg Asp Pro Lys Gly Ser Ser Ser Pro Thr Leu Phe Thr
145 150 155 160

Leu Gln Asp His Leu Lys Leu Ser Asn Met Thr Glu Ser Val Ile Gln
165 170 175

```
Asp Ser Val Ser Ile Gln Ile Cys Tyr Gln Met Arg Ile Leu Gly Phe
            180                 185                 190

Gln Cys His Ile Arg Trp Lys Leu Gln Pro Glu Asp Phe Gln Arg Asn
        195                 200                 205

Tyr Ser Pro Asn Gln Ile Gly Arg Val Val Tyr Leu Leu Tyr Glu Val
    210                 215                 220

Arg Trp Arg Arg Gly Ser Ile Trp Arg Asn Trp Cys Ser Asn Asn Pro
225                 230                 235                 240

Glu Gln His Ala Glu Val Asn Phe Leu Glu Asn His Phe His His Arg
                245                 250                 255

Pro Gln Thr Pro Cys Ser Ile Thr Trp Phe Leu Ser Thr Ser Pro Cys
            260                 265                 270

Gly Lys Cys Ser Arg Arg Ile Leu Glu Phe Leu Lys Ser Gln Pro Asn
        275                 280                 285

Val Thr Leu Glu Ile Tyr Ala Ala Lys Leu Phe Arg His His Asp Ile
    290                 295                 300

Arg Asn Arg Gln Gly Leu Arg Asn Leu Met Met Asn Gly Val Thr Ile
305                 310                 315                 320

Tyr Ile Met Asn Leu Glu Gly Asn Pro Ala Ser Leu Cys Leu Ser Val
                325                 330                 335

Asp
```

<210> 78
<211> 237
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.9;
tr|A0A3P4LUZ8|A0A3P4LUZ8_GULGU polypeptide sequence

<400> 78

```
Met Ser Phe Glu Asp Tyr Glu Tyr Cys Trp Glu Thr Phe Val Asp His
1               5                  10                  15

Lys Gly Met Tyr Phe Gln Ser Trp Asp Leu Leu Arg Asp Asn Asp Leu
            20                  25                  30

Leu Ala Ala Glu Leu Lys Asn Ile Leu Arg Ser Thr Met Asn Pro Leu
        35                  40                  45

Arg Gln Glu Ile Phe Tyr His Gln Phe Gly Asn Gln Pro Arg Ala Pro
    50                  55                  60

Arg Pro Tyr His Arg Arg Lys Thr Tyr Leu Cys Tyr Gln Leu Gln Pro
65                  70                  75                  80

His Glu Gly Pro Ile Thr Ala Arg Val Cys Leu Gln Asn Lys Lys Lys
                85                  90                  95

Arg His Ala Glu Ile Arg Phe Ile Asp Asn Ile Arg Ala Leu Arg Leu
            100                 105                 110

Asp Arg Ser Gln Thr Phe Glu Ile Thr Cys Tyr Leu Thr Trp Ser Pro
        115                 120                 125

Cys Pro Thr Cys Ala Lys Ala Leu Ala Val Phe Val Gln Asp His Pro
    130                 135                 140

His Ile Ser Leu Arg Leu Phe Ala Ser Arg Leu Phe Ile His Trp Cys
145                 150                 155                 160

Trp Lys Tyr Gln Glu Gly Leu Arg Leu Leu His Arg Ser Arg Ile Pro
                165                 170                 175

Val Ala Val Met Arg Leu Gln Glu Phe Glu Asp Cys Trp Arg Asn Phe
            180                 185                 190

Val Asp Asn Gln Asp Glu Pro Phe Gln Pro Trp Asn Lys Leu Glu Gln
        195                 200                 205

Tyr Ser Glu Ser Ile Thr Arg Arg Leu Arg Arg Ile Leu Gly His Pro
```

210 215 220

Gln Asn Asn Leu Glu Asn Asp Phe Arg Asn Leu His Ile
225 230 235

<210> 79
<211> 221
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.10;
tr|G5BPM8|G5BPM8_HETGA polypeptide sequence

<400> 79

Arg Arg Arg Ile Glu Pro Trp Gln Phe Glu Ala Ser Phe Asp Pro Arg
1 5 10 15

Gln Leu Arg Arg Glu Thr Cys Leu Leu Ser Glu Val Arg Trp Gly Thr
20 25 30

Ser Pro Arg Ala Trp Arg Gly Cys Ser Leu Asn Thr Ala Arg His Ala
35 40 45

Glu Val Ser Phe Met Asp Arg Leu Thr Ser Glu Gly Arg Leu Arg Gly
50 55 60

Pro Val Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly
65 70 75 80

Ala Cys Ala Gln Ala Ile Gly Glu Phe Leu Arg Gln His Pro Asn Val
85 90 95

Ser Leu Val Ile Tyr Ile Ala Arg Leu Phe Trp His Val Asp Glu Gln
100 105 110

Asn Arg Gln Gly Leu Arg Asp Leu Val Thr Arg Gly Val Arg Met Gln
115 120 125

Val Met Ser Asp Pro Glu Phe Ala His Cys Trp Arg Asn Phe Val Asn
130 135 140

```
Tyr Ser Pro Gly Gln Glu Ala Arg Trp Pro Gln Val Pro Pro Val Trp
145                 150                 155                 160


Thr Trp Leu Tyr Ser Leu Glu Leu His Cys Ile Leu Leu Asn Leu Pro
                165                 170                 175


Pro Cys Leu Lys Ile Ser Arg Arg His His Asn Gln Leu Thr Phe Phe
            180                 185                 190


Gln Leu Ile Leu Gln Asn Cys His Tyr Gln Ala Ile Pro Ser Pro Val
        195                 200                 205


Leu Leu Ala Ser Gly Leu Ile His Pro Phe Val Thr Trp
    210                 215                 220


<210>  80
<211>  211
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.11;
       tr|H2M862|H2M862_ORYLA polypeptide sequence

<400>  80

Met Ile Thr Lys Leu Asp Ser Val Leu Leu Pro Lys Lys Lys Phe Ile
1               5                   10                  15


Tyr His Tyr Lys Asn Met Arg Trp Ala Arg Gly Arg His Glu Thr Tyr
            20                  25                  30


Leu Cys Phe Val Val Lys Arg Arg Val Gly Pro Glu Ser Leu Ser Phe
        35                  40                  45


Asp Phe Gly His Leu Arg Asn Arg Asn Gly Cys His Val Glu Leu Leu
    50                  55                  60


Phe Leu Arg His Leu Ser Ala Leu Cys Pro Gly Leu Trp Gly Tyr Gly
65                  70                  75                  80


Ala Thr Gly Gln Gly Arg Val Ser Tyr Ser Ile Thr Trp Phe Cys Ser
                85                  90                  95
```

```
Trp Ser Pro Cys Ala Asn Cys Ser Phe Arg Leu Ala Gln Phe Leu Ser
            100                 105                 110

Gln Thr Pro Asn Leu Arg Leu Arg Ile Phe Val Ser Arg Leu Tyr Phe
        115                 120                 125

Cys Asp Leu Glu Asp Ser Arg Glu Arg Glu Gly Leu Arg Met Leu Lys
    130                 135                 140

Lys Val Gly Val His Ile Thr Val Met Ser Tyr Lys Asp Tyr Phe Tyr
145                 150                 155                 160

Cys Trp Gln Thr Phe Val Ala Arg Lys Gln Ser Lys Phe Lys Pro Trp
                165                 170                 175

Asp Gly Leu His Gln Asn Ser Val Arg Leu Ser Arg Lys Leu Asn Arg
            180                 185                 190

Ile Leu Gln Pro Cys Glu Thr Glu Asp Phe Arg Asp Ala Phe Lys Leu
        195                 200                 205

Leu Gly Leu
    210
```

<210> 81
<211> 353
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.12;
tr|H0Y0C6|H0Y0C6_OTOGA polypeptide sequence

<400> 81

```
Met Tyr Leu Lys Thr Phe Tyr Arg His Phe Asn Asn Arg Pro Tyr Leu
1               5                   10                  15

Ser Arg Arg Asn Asp Thr Trp Leu Cys Phe Glu Val Lys Thr Thr Ser
            20                  25                  30

Ser Asn Ser Pro Gly Ser Phe Tyr Ser Gly Val Phe Arg Asn Gln Gly
        35                  40                  45
```

```
Pro Arg Tyr Cys Pro Trp His Thr Glu Leu Cys Phe Leu Thr Trp Val
    50                  55                  60

Arg Pro Ile Val Ser His His His Phe Tyr Gln Ile Thr Trp Tyr Met
65                  70                  75                  80

Ser Trp Ser Pro Cys Ala Asn Cys Ala Trp Gln Val Ala Thr Phe Leu
                85                  90                  95

Ala Thr His Glu Asn Val Ser Leu Thr Asn Tyr Thr Val Arg Ile Tyr
            100                 105                 110

Tyr Phe Trp Arg Gln Asp Tyr Arg Gln Gly Leu Leu Arg Met Ile Glu
        115                 120                 125

Glu Gly Thr Gln Val Tyr Val Met Ser Ser Lys Glu Phe Gln His Cys
    130                 135                 140

Trp Glu Asn Phe Val Asp His Trp Gly Thr Arg Trp Val Thr Cys Trp
145                 150                 155                 160

Asn Arg Leu Lys Lys Asn Tyr Glu Phe Leu Val Thr Arg Leu Ser Glu
                165                 170                 175

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

Gln Phe Asn Asn Thr Pro Val Pro Arg Gly Arg Lys Asp Thr Trp Leu
        195                 200                 205

Cys Phe Glu Val Lys Glu Lys Asn Ser Asn Ser Pro Gly Ser Phe His
    210                 215                 220

Arg Gly Val Phe Gln Asn Gln Val Phe Ser Gly Thr Ser Ser His Ala
225                 230                 235                 240

Arg Arg Cys Pro Pro Asp His His Tyr Glu Val Thr Trp Tyr Thr Ser
                245                 250                 255
```

```
Trp Ser Pro Cys Ala His Cys Ala Trp His Val Val Asn Phe Leu Thr
            260                 265                 270

Ser Asn Pro Asn Val Ser Leu Thr Ile Phe Ala Ala Arg Leu Tyr Tyr
        275                 280                 285

Ile Tyr Arg Pro Glu Ile Gln Gln Gly Leu Arg Arg Val Phe Gln Glu
    290                 295                 300

Gly Ala Lys Val His Ile Met Ser Leu Lys Glu Phe Lys Tyr Cys Trp
305                 310                 315                 320

Ala Lys Leu Val Tyr Asn Ser Gly Met Arg Phe Met Pro Trp Tyr Gln
                325                 330                 335

Phe Asn Phe Asn Phe Leu Phe Pro Asn Thr Thr Leu Lys Gly Asp Leu
            340                 345                 350

His
```

<210> 82
<211> 189
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.13;
tr|A0A3Q2Z5X6|A0A3Q2Z5X6_HIPCM polypeptide sequence

<400> 82

```
Met Asp Val His Phe Met Asn Phe Ile Tyr His Tyr Lys Asn Met Arg
1               5                   10                  15

Trp Ala Lys Gly Arg Asn Glu Thr Tyr Leu Cys Phe Val Val Lys Arg
            20                  25                  30

Arg Val Gly Pro Asn Ser Leu Thr Phe Asp Phe Gly His Leu Arg Asn
        35                  40                  45

Arg Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr Leu Gly Arg
    50                  55                  60
```

```
Arg Leu Ser Tyr Ser Ile Thr Trp Phe Cys Ser Trp Ser Pro Cys Ala
65                  70                  75                  80


Asn Cys Ser Ala Ala Leu Ser Gln Phe Leu Ser Arg Met Pro Asn Leu
                85                  90                  95


Arg Leu Arg Ile Phe Val Ala Arg Leu Tyr Phe Cys Asp Met Glu Asp
            100                 105                 110


Ser His Glu Arg Glu Gly Leu Arg Leu Leu Gln Lys Ala Gly Val Gln
        115                 120                 125


Val Thr Val Met Ser Tyr Lys Asp Tyr Tyr Tyr Cys Trp Gln Thr Phe
    130                 135                 140


Val Asp Arg Lys Lys Ser His Phe Lys Ala Trp Glu Asp Leu His Gln
145                 150                 155                 160


Asn Ser Val Arg Leu Ser Arg Lys Leu Asn Arg Ile Leu Gln Pro Cys
                165                 170                 175


Glu Met Asp Leu Arg Asp Ala Phe Lys Leu Leu Gly Leu
            180                 185


<210>  83
<211>  190
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.15;
       tr|A0A2K6NY90|A0A2K6NY90_RHIRO polypeptide sequence

<400>  83

Met Asn Pro His Ile Arg Asn Pro Met Glu Ala Met Tyr Pro Gly Thr
1               5                   10                  15


Phe Tyr Phe His Phe Lys Asn Leu Trp Glu Ala Asp Asn Arg Asn Glu
            20                  25                  30


Ser Trp Leu Cys Phe Ala Val Glu Val Ile Lys His His Ser Thr Val
        35                  40                  45
```

```
Ser Trp Lys Arg Gly Val Phe Arg Asn Gln Val Asp Pro Glu Thr His
    50                  55                  60

Cys His Ala Glu Lys Cys Phe Leu Ser Trp Phe Cys Asp Asn Thr Leu
65                  70                  75                  80

Ser Pro Lys Lys Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
                85                  90                  95

Cys Pro Glu Cys Ala Arg Glu Val Ala Lys Phe Leu Ala Arg His Ser
            100                 105                 110

Asn Val Met Leu Thr Ile Tyr Thr Ala Arg Leu Tyr Tyr Ser Gln Tyr
        115                 120                 125

Pro Asn Tyr Gln Glu Gly Leu Arg Arg Leu Asn Glu Glu Gly Val Pro
    130                 135                 140

Val Glu Ile Met Asp Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145                 150                 155                 160

Val Tyr Asn Gly Asp Glu Leu Phe Lys Pro Trp Lys Gly Leu Lys Tyr
                165                 170                 175

Asn Phe Leu Phe Leu Asp Ser Lys Leu Gln Glu Ile Leu Glu
            180                 185                 190
```

<210> 84
<211> 202
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.16;
       tr|Q6ICH2|Q6ICH2_HUMAN polypeptide sequence

<400> 84

```
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1               5                   10                  15

Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
```

```
            20                  25                  30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
        35                  40                  45

Leu Trp Asp Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln
    50                  55                  60

Ser Asn His Arg Gln Glu Val Asp Pro Glu Thr His Cys His Ala Glu
65                  70                  75                  80

Arg Cys Phe Leu Ser Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr
                85                  90                  95

Asn Tyr Glu Val Thr Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys
            100                 105                 110

Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu
        115                 120                 125

Thr Ile Phe Thr Ala Arg Leu Cys Tyr Phe Trp Asp Thr Asp Tyr Gln
    130                 135                 140

Glu Gly Leu Cys Ser Leu Ser Gln Glu Gly Ala Ser Val Lys Ile Met
145                 150                 155                 160

Gly Tyr Lys Asp Phe Val Ser Cys Trp Lys Asn Phe Val Tyr Ser Asp
                165                 170                 175

Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Gln Thr Asn Phe Arg Leu
            180                 185                 190

Leu Lys Arg Arg Leu Arg Glu Ile Leu Gln
        195                 200
```

<210> 85
<211> 202
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.17;

tr|G8GPV1|G8GPV1_CERNE polypeptide sequence

<400> 85

```
Met Asp Gly Ser Pro Ala Ser Arg Pro Gly His Val Met Asp Pro Gly
1               5                  10                  15

Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln Arg
            20                  25                  30

Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Ser His Asn Asp Thr Trp
        35                  40                  45

Val Leu Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Lys Asn
    50                  55                  60

Arg Leu His Gly Asp Tyr Gly Cys His Ala Glu Leu Cys Phe Leu Gly
65                  70                  75                  80

Glu Val Pro Ser Trp Arg Leu Asp Pro Thr Gln Thr Tyr Arg Val Thr
                85                  90                  95

Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Gly Gly Cys Ala Glu Gln
            100                 105                 110

Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe
        115                 120                 125

Ala Ala Arg Ile Tyr Asp Tyr Asp Phe Leu Tyr Gln Glu Ala Leu Arg
    130                 135                 140

Thr Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Glu Glu
145                 150                 155                 160

Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Arg Pro Phe
                165                 170                 175

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg
            180                 185                 190

Leu Gln Ala Ile Leu Gln Asn Gln Gly Asn
        195                 200
```

<210> 86
<211> 181
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.18;
tr|Q1WBT6|Q1WBT6_SYMSY polypeptide sequence

<400> 86

Met Ala Leu Leu Thr Ala Lys Thr Phe Arg Leu Gln Phe Asn Asn Lys
1 5 10 15

Arg Arg Val Thr Lys Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr
20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Lys
35 40 45

Asn Lys Lys Lys Arg His Ala Glu Ile Arg Phe Ile Asn Lys Ile Lys
50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu
65 70 75 80

Thr Trp Ser Pro Cys Pro Ser Cys Ala Trp Glu Leu Val Asp Phe Ile
85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr
100 105 110

Tyr His Trp Cys Arg His Gln Gln Glu Gly Leu Arg Leu Leu Cys Gly
115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Glu Phe Ala Asp Cys
130 135 140

Trp Glu Asn Phe Val Asp His Glu Glu Pro Leu Ser Phe Asn Pro Ser
145 150 155 160

Glu Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg

```
                165                 170                 175


Leu Glu Lys Ile Lys
            180


<210>  87
<211>  211
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.19;
       tr|A0A3B4CS14|A0A3B4CS14_PYGNA polypeptide sequence

<400>  87

Met Asp Asn Thr Asn Arg Arg Lys Phe Ile Tyr His Tyr Lys Asn Val
1               5                   10                  15


Arg Trp Ala Arg Gly Arg His Glu Thr Tyr Leu Cys Phe Val Val Lys
            20                  25                  30


Lys Arg Asn Ser Pro Asp Ser Leu Ser Phe Asp Phe Gly His Leu Arg
        35                  40                  45


Asn Arg Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr Ile Glu
    50                  55                  60


Val Leu Cys Pro Gly Leu Trp Gly Ser Gly Val Asp Gly Val Arg Val
65                  70                  75                  80


Ser Tyr Ala Val Thr Trp Phe Cys Ser Trp Ser Pro Cys Ser Asn Cys
                85                  90                  95


Ala Gln Arg Leu Thr Asn Phe Leu Ser Gln Thr Pro Asn Leu Arg Leu
            100                 105                 110


Arg Ile Phe Val Ala Arg Leu Tyr Phe Cys Asp Glu Glu Asp Ser Leu
        115                 120                 125


Glu Arg Glu Gly Leu Arg His Leu Gln Arg Ala Gly Val Gln Ile Thr
    130                 135                 140
```

```
Val Met Thr Tyr Lys Asp Phe Phe Tyr Cys Trp Gln Thr Phe Val Ala
145                 150                 155                 160


Ser Arg Glu Arg Cys Phe Lys Ala Trp Glu Gly Leu Arg Gln Asn Ser
                165                 170                 175


Val Arg Leu Ser Arg Lys Leu Asn Arg Ile Leu Gln Val Phe Ile Ser
            180                 185                 190


Thr Pro Val Ile Ser Pro Leu Ile Thr Thr His Leu Gly Gln Ser Trp
        195                 200                 205


Ala Gly Gly
    210
```

<210> 88
<211> 315
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.20;
       tr|A0A087XZI4|A0A087XZI4_POEFO polypeptide sequence

<400> 88

```
Arg Lys Val Ser Tyr Ser Val Thr Trp Phe Cys Ser Trp Ser Pro Cys
1               5                   10                  15


Ala Asn Cys Ser Ile Arg Leu Ala Gln Phe Leu His Gln Thr Pro Asn
            20                  25                  30


Leu Arg Leu Arg Ile Phe Val Ser Arg Leu Tyr Phe Cys Asp Leu Glu
        35                  40                  45


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


His Ile Thr Val Met Ser Tyr Lys Asp Tyr Phe Tyr Cys Trp Gln Thr
65                  70                  75                  80


Phe Val Ala Lys Ser Gln Ser Lys Phe Lys Pro Trp Asp Gly Leu His
                85                  90                  95
```

```
Gln Asn Tyr Ile Arg Leu Ser Arg Lys Leu Asn Arg Ile Leu Gln Pro
            100                 105                 110

Ala Leu Asp Ile Lys Lys Phe Ile Tyr His Tyr Lys Asn Leu Arg Trp
        115                 120                 125

Ala Arg Gly Arg Cys Glu Thr Tyr Leu Cys Phe Val Val Lys Lys Lys
    130                 135                 140

Leu His Leu Phe Met Phe Val Ile Val Gly Arg Asn Arg Leu Phe Asp
145                 150                 155                 160

Leu Asn Val Thr Met Asn Asn Lys Ser Leu Tyr Leu Ile Pro Leu His
                165                 170                 175

Leu Gln Leu Leu Phe Leu Arg His Leu Gly Ala Leu Cys Pro Gly Leu
            180                 185                 190

Trp Gly Tyr Gly Val Thr Gly Glu Arg Lys Val Ser Tyr Ser Val Thr
        195                 200                 205

Trp Phe Cys Ser Trp Ser Pro Cys Ala Asn Cys Ser Ile Arg Leu Ala
    210                 215                 220

Gln Phe Leu His Gln Thr Pro Asn Leu Arg Leu Arg Ile Phe Val Ser
225                 230                 235                 240

Arg Leu Tyr Phe Cys Asp Leu Glu Asp Ser Arg Glu Arg Glu Gly Leu
                245                 250                 255

Arg Ile Leu Lys Lys Ala Gly Val His Ile Thr Val Met Ser Tyr Lys
            260                 265                 270

Asp Tyr Phe Tyr Cys Trp Gln Thr Phe Val Ala Lys Ser Gln Ser Lys
        275                 280                 285

Phe Lys Pro Trp Asp Gly Leu His Gln Asn Tyr Ile Arg Leu Ser Arg
    290                 295                 300

Lys Leu Asn Arg Ile Leu Gln Val Gln Phe Phe
```

305 310 315

<210> 89
<211> 189
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.21;
tr|A0A341AEK4|A0A341AEK4_9CETA polypeptide sequence

<400> 89

Met Ala Ser Asp Arg Gly Pro Ser Ala Gly Asp Ala Thr Ser Arg Arg
1 5 10 15

Arg Ile Glu Pro Trp Glu Phe Glu Val Ser Phe Asp Pro Arg Glu Leu
20 25 30

Cys Lys Glu Thr Arg Leu Leu Tyr Glu Ile Lys Trp Gly Arg Ser Gln
35 40 45

His Val Trp Arg His Ser Gly Lys Asn Thr Thr Asn His Val Glu Cys
50 55 60

Asn Phe Ile Glu Lys Phe Thr Ser Glu Arg Pro Phe His Arg Ser Val
65 70 75 80

Ser Cys Cys Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95

Ser Lys Ala Ile Arg Glu Phe Leu Asn Gln His Pro Arg Val Thr Leu
100 105 110

Phe Ile Tyr Val Ala Arg Leu Phe Gln His Met Asp Pro Gln Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile His Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Gly Pro Thr Glu Tyr Asp Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160

Pro Gly Lys Glu Ala His Trp Pro Arg Tyr Pro Pro Pro Leu Met Lys
165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Val Pro
180 185

<210> 90
<211> 180
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.22;
tr|E2D879|E2D879_MUSMI polypeptide sequence

<400> 90

Arg Asn Leu Ile Ser Arg Glu Thr Phe Asn Phe Asn Phe Glu Asn Leu
1 5 10 15

Cys Tyr Ala Lys Gly Arg Lys Asn Thr Phe Leu Cys Tyr Glu Val Thr
20 25 30

Arg Lys Asp Cys Asp Ser Pro Val Ser Leu Cys His Gly Val Phe Lys
35 40 45

Asn Lys Gly Ser Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe His
50 55 60

Asp Lys Val Leu Lys Val Leu Thr Pro Arg Glu Glu Phe Lys Val Thr
65 70 75 80

Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Val Val
85 90 95

Arg Phe Leu Ala Thr His His Asn Leu Asn Leu Thr Ile Phe Ser Ser
100 105 110

Arg Leu Tyr Asn Val Ser Asp Pro Asp Thr Gln Gln Lys Leu Cys Arg
115 120 125

Leu Val Gln Glu Gly Ala Gln Val Ala Val Met Asp Leu Ser Glu Phe
130 135 140

```
Lys Lys Cys Trp Glu Lys Phe Val Asp Asn Asp Gly Gln Gln Phe Arg
145                 150                 155                 160


Pro Trp Lys Arg Leu Arg Thr Asn Phe Arg Tyr Gln Asn Ser Lys Leu
                165                 170                 175


Gln Glu Ile Leu
            180


<210>  91
<211>  221
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.23;
       tr|A0A2K5RDN6|A0A2K5RDN6_CEBCA polypeptide sequence

<400>  91

Met Trp Glu Ala Gln Ser Pro Gly Leu Ser Arg Glu Trp Gly Ser Val
1               5                   10                  15


Ala Ile Ser Pro Glu Asp Pro Gly Pro Leu His Ile Gly Arg Phe Leu
            20                  25                  30


Ser Cys Ala Phe Arg His Pro Met Asn Ala Met Tyr Pro Gly Ile Phe
        35                  40                  45


Asn Phe His Phe Arg Asn Leu Arg Lys Ala Tyr Gly Arg Asn Glu Thr
    50                  55                  60


Trp Leu Cys Phe Thr Val Glu Gly Ile Met Asn Arg Ser Thr Val Ser
65                  70                  75                  80


Trp Lys Ser Gly Val Phe Arg Asn Gln Val Gly Ser Asp Pro Phe Cys
                85                  90                  95


His Ala Glu Met Cys Phe Leu Ser Trp Phe Arg His Asn Met Leu Ser
            100                 105                 110


Pro Lys Lys Asp Tyr Glu Val Thr Trp Tyr Ala Ser Trp Ser Pro Cys
        115                 120                 125
```

```
Pro Glu Cys Ala Gly Gln Val Ala Glu Phe Leu Ala Arg His Gly Asn
    130                 135                 140

Val Arg Leu Thr Ile Phe Thr Ala His Leu Tyr Tyr Phe Trp Asn Pro
145                 150                 155                 160

Ser Phe Arg Gln Gly Leu Arg Arg Leu Ser Gln Glu Gly Ala Ser Val
                165                 170                 175

Leu Ile Met Gly Tyr Glu Asp Phe Glu Tyr Cys Trp Asp Asn Phe Val
            180                 185                 190

Tyr Asn Asp Gly Gln Pro Phe Lys Pro Trp Lys Arg Leu Gln Asp Asn
        195                 200                 205

Ser Leu Ser Leu Tyr Ile Thr Leu Gln Glu Ile Leu Gln
    210                 215                 220
```

<210> 92
<211> 206
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.24;
tr|A0A2K5RDN7|A0A2K5RDN7_CEBCA polypeptide sequence

<400> 92

```
Met Glu Ala Ser Pro Ala Ser Arg Pro Arg Pro Leu Met Gly Pro Arg
1               5                   10                  15

Thr Phe Thr Glu Asn Phe Thr Asn Asn Pro Glu Val Phe Gly Arg His
            20                  25                  30

Gln Thr Tyr Leu Cys Tyr Glu Val Lys Cys Gln Gly Pro Asp Gly Thr
        35                  40                  45

Arg Asp Leu Met Thr Glu Gln Arg Asp Phe Leu Cys Asn Gln Ala Arg
    50                  55                  60

Asn Leu Leu Ser Gly Phe Asp Gly Arg His Ala Glu Arg Cys Phe Leu
```

65 70 75 80

Asp Arg Val Pro Ser Trp Arg Leu Asp Pro Ala Gln Thr Tyr Arg Val
85 90 95

Thr Cys Phe Ile Ser Trp Ser Pro Cys Phe Ser Cys Ala Arg Glu Val
100 105 110

Ala Glu Phe Leu Gln Glu Asn Pro His Val Asn Leu Arg Ile Phe Ala
115 120 125

Ala Arg Ile Tyr Asp Cys Arg Pro Arg Tyr Glu Glu Gly Leu Gln Met
130 135 140

Leu Gln Asn Ala Gly Ala Gln Val Ser Ile Met Thr Ser Glu Glu Phe
145 150 155 160

Arg His Cys Trp Asp Thr Phe Val Asp His Gln Gly His Pro Phe Gln
165 170 175

Pro Trp Glu Gly Leu Asp Glu His Ser Gln Ala Leu Ser Arg Arg Leu
180 185 190

Gln Ala Ile Leu Gln Gly Asn Arg Trp Met Ile Leu Ser Leu
195 200 205

<210> 93
<211> 184
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.25;
tr|A0A1C9CJ69|A0A1C9CJ69_CERAL polypeptide sequence

<400> 93

Asn Pro Met Lys Ala Met Asp Pro His Ile Phe Tyr Phe His Phe Lys
1 5 10 15

Asn Leu Arg Lys Ala Tyr Gly Arg Asn Glu Thr Trp Leu Cys Phe Ala
20 25 30

```
Val Glu Ile Ile Lys Gln Arg Ser Thr Val Pro Trp Arg Thr Gly Val
        35                  40                  45

Phe Arg Asn Gln Val Asp Pro Glu Ser His Cys His Ala Glu Arg Cys
    50                  55                  60

Phe Leu Ser Trp Phe Cys Glu Asp Ile Leu Ser Pro Asn Thr Asp Tyr
65                  70                  75                  80

Arg Val Thr Trp Tyr Thr Ser Trp Ser Pro Cys Leu Asp Cys Ala Gly
                85                  90                  95

Glu Val Ala Glu Phe Leu Ala Arg His Ser Asn Val Glu Leu Ala Ile
            100                 105                 110

Phe Ala Ala Arg Leu Tyr Tyr Phe Trp Asp Thr His Tyr Gln Gln Gly
        115                 120                 125

Leu Arg Ser Leu Ser Glu Lys Gly Ala Ser Val Glu Ile Met Gly Tyr
    130                 135                 140

Glu Asp Phe Lys Tyr Cys Arg Glu Asn Phe Val Cys Asp Asp Gly Lys
145                 150                 155                 160

Pro Phe Lys Pro Trp Lys Gly Leu Lys Thr Asn Phe Arg Phe Leu Lys
                165                 170                 175

Arg Arg Leu Gln Glu Ile Leu Glu
            180
```

<210> 94
<211> 197
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.26; tr|A0A2R2Z4D2|A0A2R2Z4D2_PTEAL polypeptide sequence

<400> 94

```
Met His Leu Gln Val Trp Arg Lys Val Thr Glu Ala Trp Arg Glu Gly
1               5                   10                  15
```

```
Tyr Thr Leu Lys Pro Trp Ser Arg Asn Pro Met Glu Arg Leu Tyr His
            20                  25                  30

Asp Tyr Phe Tyr Phe His Phe Tyr Asn Leu Pro Thr Pro Lys His Arg
        35                  40                  45

Asn Gly Cys Tyr Ile Cys Tyr Gln Val Glu Gly Thr Lys Lys His Ser
    50                  55                  60

Arg Met Pro Leu Leu Arg Gly Val Phe Glu Asn Gln Glu Ser Leu Asp
65                  70                  75                  80

Met Met Leu Ser Pro Gly Glu Lys Tyr Arg Val Thr Trp Tyr Ile Ser
                85                  90                  95

Trp Ser Pro Cys Phe Ala Cys Val Asp Glu Val Ile Lys Phe Leu Arg
            100                 105                 110

Glu His Thr Asn Val Glu Leu Ile Ile Phe Ala Ala Arg Leu Tyr His
        115                 120                 125

Ser Asp Ile Leu Gln Tyr Arg Gln Gly Leu Arg Lys Leu His Asp Ala
    130                 135                 140

Gly Val His Val Ala Ile Met Ser Tyr Tyr Glu Phe Lys His Cys Leu
145                 150                 155                 160

Asn Asp Phe Val Phe His Gln Gly Arg Ser Phe Cys Pro Trp Asn Asp
                165                 170                 175

Leu Asn Lys Asn Ser Lys Asn Leu Ser Asn Thr Leu Glu Asp Ile Leu
            180                 185                 190

Arg Asn Gln Glu Asp
        195
```

<210> 95
<211> 206
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.27;
tr|B7T161|B7T161_SHEEP polypeptide sequence

<400> 95

```
Met Thr Glu Gly Trp Ala Gly Ser Gly Leu Pro Gly Arg Gly Asp Cys
1               5                  10                  15


Val Trp Thr Pro Gln Thr Arg Asn Thr Met Asn Leu Leu Arg Glu Thr
            20                  25                  30


Leu Phe Lys Gln Gln Phe Gly Asn Gln Pro Arg Val Pro Pro Pro Tyr
        35                  40                  45


Tyr Arg Arg Lys Thr Tyr Leu Cys Tyr Gln Leu Lys Glu Leu Asp Asp
    50                  55                  60


Leu Met Leu Asp Lys Gly Cys Phe Arg Asn Lys Lys Gln Arg His Ala
65                  70                  75                  80


Glu Ile Arg Phe Ile Asp Lys Ile Asn Ser Leu Asn Leu Asn Pro Ser
                85                  90                  95


Gln Ser Tyr Lys Ile Ile Cys Tyr Ile Thr Trp Ser Pro Cys Pro Asn
            100                 105                 110


Cys Ala Ser Glu Leu Val Asp Phe Ile Thr Arg Asn Asp His Leu Asn
        115                 120                 125


Leu Gln Ile Phe Ala Ser Arg Leu Tyr Phe His Trp Ile Lys Pro Phe
    130                 135                 140


Cys Arg Gly Leu His Gln Leu Gln Lys Ala Gly Ile Ser Val Ala Val
145                 150                 155                 160


Met Thr His Thr Glu Phe Glu Asp Cys Trp Glu Gln Phe Val Asp Asn
                165                 170                 175


Gln Leu Arg Pro Phe Gln Pro Trp Asp Lys Leu Glu Gln Tyr Ser Ala
            180                 185                 190
```

Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr Ala Pro Thr
195 200 205

<210> 96
<211> 205
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.28; tr|A0A2R2X2G4|A0A2R2X2G4_PTEAL polypeptide sequence

<400> 96

Met Ala Gly Leu Gly Gln Ala Cys Glu Gly Cys Cys Gly Gln Met Pro
1 5 10 15

Glu Ile Ser Tyr Pro Met Gly Arg Leu Asp Pro Lys Thr Phe Ser Phe
20 25 30

Glu Phe Lys Asn Leu Pro Tyr Ala Tyr Gly Arg Lys Ser Ser Tyr Leu
35 40 45

Cys Phe Gln Val Glu Arg Glu Gln His Ser Ser Pro Val Pro Ser Asp
50 55 60

Trp Gly Val Phe Lys Asn Gln Phe Cys Gly Thr Glu Pro Tyr His Ala
65 70 75 80

Glu Leu Cys Phe Leu Asn Trp Phe Arg Ala Glu Lys Leu Ser Pro Tyr
85 90 95

Glu His Tyr Asp Val Thr Trp Phe Leu Ser Trp Ser Pro Cys Ser Thr
100 105 110

Cys Ala Glu Glu Ile Ala Ile Phe Leu Ser Asn His Lys Asn Val Arg
115 120 125

Leu Asn Ile Phe Val Ser Arg Ile Tyr Tyr Phe Trp Lys Pro Ala Phe
130 135 140

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

Met Ser Phe Asp Glu Phe Lys Tyr Cys Trp Glu Asn Phe Val Asp Asn
165 170 175

Gln Gly Met Pro Phe Arg Cys Trp Lys Lys Val His Gln Asn Tyr Lys
180 185 190

Ser Val Leu Arg Lys Leu Asn Glu Ile Leu Arg Arg Arg
195 200 205

<210> 97
<211> 309
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.29;
tr|G1Q1M4|G1Q1M4_MYOLU polypeptide sequence

<400> 97

Tyr Ala Glu Leu Ser Phe Leu Asp Leu Phe Gln Ser Trp Asn Leu Asp
1 5 10 15

Arg Gly Arg Gln Tyr Arg Leu Thr Trp Tyr Met Ser Trp Ser Pro Tyr
20 25 30

Pro Asp Cys Ala Gln Lys Leu Val Glu Phe Leu Gly Glu Asn Ser His
35 40 45

Val Thr Leu Arg Ile Phe Ala Ala Asp Ile His Ser Leu Cys Ser Gly
50 55 60

Tyr Glu Asp Gly Leu Arg Lys Leu Arg Asp Ala Arg Ala Gln Leu Ala
65 70 75 80

Ile Met Thr Arg Asp Glu Leu Gln Tyr Cys Trp Val Thr Phe Val Asp
85 90 95

Asn Gln Gly Gln Pro Phe Arg Pro Trp Pro Asn Leu Val Glu His Ile
100 105 110

Lys Thr Lys Lys Gln Glu Leu Lys Asp Ile Leu Gly Asn Pro Met Arg
115 120 125

```
Arg Met Tyr Pro Lys Thr Phe Asn Phe Asn Phe Gln Asn Leu Asn Ser
    130                 135                 140

Tyr Gly Arg Lys Ser Thr Phe Leu Cys Phe Glu Val Glu Thr Trp Glu
145                 150                 155                 160

Asp Gly Ser Val Leu Asp Tyr Gln Asn Gly Val Phe Gln Asn Gln Leu
                165                 170                 175

Asp Pro Gly His Ala Glu Leu Cys Phe Ile Glu Trp Phe His Glu Lys
            180                 185                 190

Val Leu Phe Pro Asp Glu Val Arg Cys Pro Asp Ala Gln Tyr His Val
        195                 200                 205

Thr Trp Tyr Ile Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Val
    210                 215                 220

Ala Gly Phe Leu Asn Glu His Glu Asn Val Asp Leu Ser Ile Ser Ala
225                 230                 235                 240

Ala Arg Leu Tyr Leu Cys Glu Asp Glu Asp Glu Gln Gly Leu Gln Asp
                245                 250                 255

Leu Val Ala Ala Gly Ala Lys Val Ala Met Met Ala Pro Glu Asp Phe
            260                 265                 270

Glu Tyr Cys Trp Asp Asn Phe Val Tyr Asn Arg Gly Trp Pro Phe Thr
        275                 280                 285

Tyr Trp Lys His Val Arg Arg Asn Tyr Gly Arg Leu Gln Glu Lys Leu
    290                 295                 300

Asp Glu Ile Leu Trp
305
```

<210> 98
<211> 216
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.30;
tr|A0A1S3AN78|A0A1S3AN78_ERIEU polypeptide sequence

<400> 98

Arg Arg Ile Glu Pro Trp Glu Phe Glu Asp Phe Phe Asp Pro Arg Gln
1 5 10 15

Phe Arg Pro Glu Thr Cys Leu Leu Tyr Glu Val Arg Trp Gly Ser Ser
20 25 30

Arg Asn Ala Trp Arg Ser Thr Ala Arg Asn Thr Thr Arg His Ala Glu
35 40 45

Val Asn Phe Leu Glu Arg Phe Ala Ala Glu Arg His Phe Asp Lys Pro
50 55 60

Val Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu
65 70 75 80

Cys Ser Gln Ala Ile Gly Ala Phe Leu Ser Gln His Pro Gln Val Thr
85 90 95

Leu Ala Ile His Val Thr Arg Leu Phe His His Glu Asp Glu Gln Asn
100 105 110

Arg Gln Gly Leu Arg Asp Leu Leu Ala Arg Gly Val Thr Leu Gln Val
115 120 125

Met Gly Asp Ser Glu Tyr Ala His Cys Trp Arg Thr Phe Val Asn Ser
130 135 140

Pro Pro Gly Ala Glu Gly His Tyr Pro Arg Tyr Pro Ser Asp Phe Thr
145 150 155 160

Arg Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Gly Leu Pro Pro
165 170 175

Cys Leu Glu Ile Leu Arg Arg Tyr Gln Asn Gln Phe Thr Leu Phe Arg
180 185 190

Leu Val Pro Gln Asn Cys His Tyr Gln Met Ile Pro His Leu Asn Phe
195 200 205

Phe Val Val Arg His Tyr Phe Phe
210 215

<210> 99
<211> 201
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.32;
tr|Q4VUI3|Q4VUI3_XENLA polypeptide sequence

<400> 99

Met Thr Met Asp Ser Met Leu Leu Lys Arg Asn Lys Phe Ile Tyr His
1 5 10 15

Tyr Lys Asn Leu Arg Trp Ala Arg Gly Arg His Glu Thr Tyr Leu Cys
20 25 30

Tyr Ile Val Lys Arg Arg Tyr Ser Ser Val Ser Cys Ala Leu Asp Phe
35 40 45

Gly Tyr Leu Arg Asn Arg Asn Gly Cys His Ala Glu Met Leu Phe Leu
50 55 60

Arg Tyr Leu Ser Ile Trp Val Gly His Asp Pro His Arg Asn Tyr Arg
65 70 75 80

Val Thr Trp Phe Ser Ser Trp Ser Pro Cys Tyr Asp Cys Ala Lys Arg
85 90 95

Thr Leu Glu Phe Leu Lys Gly His Pro Asn Phe Ser Leu Arg Ile Phe
100 105 110

Ser Ala Arg Leu Tyr Phe Cys Glu Glu Arg Asn Ala Glu Pro Glu Gly
115 120 125

Leu Arg Lys Leu Gln Lys Ala Gly Val Arg Leu Ser Val Met Ser Tyr
130 135 140

```
Lys Asp Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Thr Arg Glu Ser
145                 150                 155                 160


Gly Phe Glu Ala Trp Asp Gly Leu His Glu Asn Ser Val Arg Leu Ala
                165                 170                 175


Arg Lys Leu Arg Arg Ile Leu Gln Pro Pro Tyr Asp Met Glu Asp Leu
            180                 185                 190


Arg Glu Val Phe Val Leu Leu Gly Leu
        195                 200


<210>  100
<211>  191
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.33;
       tr|E2RL86|E2RL86_CANLF polypeptide sequence

<400>  100

Met Asn Pro Leu Gln Glu Glu Thr Phe Tyr Gln Gln Phe Ser Asn Gln
1               5                   10                  15


Arg Val Pro Lys Pro Thr Tyr Gln Arg Arg Thr Tyr Leu Cys Tyr Gln
            20                  25                  30


Leu Lys Pro His Glu Gly Ser Val Ile Ala Lys Val Cys Leu Gln Asn
        35                  40                  45


Gln Glu Lys Arg His Ala Glu Ile Cys Phe Ile Asp Asp Ile Lys Ser
    50                  55                  60


Arg Gln Leu Asp Pro Ser Gln Lys Phe Glu Ile Thr Cys Tyr Val Thr
65                  70                  75                  80


Trp Ser Pro Cys Pro Thr Cys Ala Lys Lys Leu Ile Ala Phe Val Asn
                85                  90                  95


Asp His Pro His Ile Ser Leu Arg Leu Phe Ala Ser Arg Leu Tyr Phe
```

```
            100                 105                 110

His Trp Arg Gln Lys Tyr Lys Arg Glu Leu Arg His Leu Gln Lys Ser
        115                 120                 125

Gly Ile Pro Leu Ala Val Met Ser Tyr Leu Glu Phe Lys Asp Cys Trp
    130                 135                 140

Glu Lys Phe Val Asp His Lys Gly Arg Pro Phe Gln Pro Trp Asn Lys
145                 150                 155                 160

Leu Lys Gln Tyr Ser Glu Ser Ile Gly Arg Arg Leu Gln Arg Ile Leu
                165                 170                 175

Gln Pro Leu Asn Asn Leu Glu Asn Asp Phe Arg Asn Leu Arg Leu
            180                 185                 190


<210>  101
<211>  226
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.34;
       tr|G1LWB0|G1LWB0_AILME polypeptide sequence

<400>  101

Ser Ser Ala Ala Pro Ala Ser Ile His Leu Leu Asp Glu Asp Thr Phe
1               5                   10                  15

Thr Glu Asn Phe Arg Asn Asp Asp Trp Pro Ser Arg Thr Tyr Leu Cys
            20                  25                  30

Tyr Lys Val Glu Gly Pro Asp Gln Gly Ser Gly Val Pro Leu Gly Gln
        35                  40                  45

Asp Lys Gly Ile Leu His Asn Lys Pro Ala Gln Gly Pro Glu Pro Ser
    50                  55                  60

Arg His Ala Glu Cys Tyr Leu Leu Glu Gln Ile Gln Ser Trp Asn Leu
65                  70                  75                  80
```

```
Asp Pro Lys Leu His Tyr Gly Val Thr Cys Phe Leu Ser Trp Ser Pro
                85                  90                  95

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

His Val Ser Leu Lys Leu Phe Ala Ser Arg Leu Tyr Thr Arg Glu Arg
        115                 120                 125

Trp Asp Glu Asp Tyr Lys Glu Gly Leu Arg Thr Leu Lys Arg Ala Gly
    130                 135                 140

Ala Ser Ile Ala Ile Met Thr Tyr Arg Glu Phe Glu His Cys Trp Lys
145                 150                 155                 160

Thr Phe Val Leu His Asp Gln Glu Gly Ser Cys Phe Gln Pro Trp Pro
                165                 170                 175

Phe Leu His Lys Glu Ser Gln Lys Phe Ser Glu Lys Leu Gln Ala Ile
            180                 185                 190

Leu Gln Val Gly Val Leu Leu Leu Ser Leu Pro Pro Pro Leu Pro Ser
        195                 200                 205

Ser Pro Leu Ser Ser Pro Trp Pro Phe Pro Ala Pro Leu Arg Ala Ser
    210                 215                 220

Thr Gly
225
```

<210> 102
<211> 238
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.35; tr|A0A1U7S7K7|A0A1U7S7K7_ALLSI polypeptide sequence

<400> 102

```
Met Gly Glu His Trp Gln Tyr Ala Gly Ser Gly Glu Tyr Ile Pro Gln
1               5                   10                  15
```

```
Asp Gln Phe Glu Glu Asn Phe Asp Pro Ser Val Leu Leu Ala Glu Thr
            20                  25                  30

His Leu Leu Ser Glu Leu Thr Trp Gly Gly Arg Pro Tyr Lys His Trp
        35                  40                  45

Tyr Glu Asn Thr Glu His Cys His Ala Glu Ile His Phe Leu Glu Asn
    50                  55                  60

Phe Ser Ser Lys Asn Arg Ser Cys Thr Ile Thr Trp Tyr Leu Ser Trp
65                  70                  75                  80

Ser Pro Cys Ala Glu Cys Ser Ala Arg Ile Ala Asp Phe Met Gln Glu
                85                  90                  95

Asn Thr Asn Val Lys Leu Asn Ile His Val Ala Arg Leu Tyr Leu His
            100                 105                 110

Asp Asp Glu His Thr Arg Gln Gly Leu Arg Tyr Leu Met Lys Met Lys
        115                 120                 125

Arg Val Thr Ile Gln Val Met Thr Ile Pro Asp Tyr Thr Tyr Cys Trp
    130                 135                 140

Asn Thr Phe Leu Glu Asp Asp Gly Glu Asp Glu Ser Asp Asp Tyr Gly
145                 150                 155                 160

Gly Tyr Ala Gly Val His Glu Asp Glu Asp Glu Ser Asp Asp Asp Asp
                165                 170                 175

Tyr Leu Pro Thr His Phe Ala Pro Trp Ile Met Leu Tyr Ser Leu Glu
            180                 185                 190

Leu Ser Cys Ile Leu Gln Gly Phe Ala Pro Cys Leu Lys Ile Ile Gln
        195                 200                 205

Gly Asn His Met Ser Pro Thr Phe Gln Leu His Val Gln Asp Gln Glu
    210                 215                 220

Gln Lys Arg Leu Leu Glu Pro Ala Asn Pro Trp Gly Ala Asp
```

225 230 235

<210> 103
<211> 220
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.36;
tr|A0A2R2X2J8|A0A2R2X2J8_PTEVA polypeptide sequence

<400> 103

Met Pro Arg Ile Gly Asn Met Asn Leu Leu Ser Glu Lys Thr Phe Asn
1 5 10 15

Tyr His Phe Gly Asn Gln Leu Arg Val Lys Lys Pro Gln Gly Arg Arg
20 25 30

Arg Thr Tyr Leu Cys Tyr Lys Leu Lys Leu Pro Asn Glu Thr Leu Val
35 40 45

Lys Gly Tyr Phe Ile Asn Lys Lys Lys Asn His Ala Glu Ile Arg Phe
50 55 60

Ile Asn Lys Ile Arg Ser Leu Asn Leu Asp Gln Thr Gln Ser Tyr Lys
65 70 75 80

Ile Thr Cys Tyr Ile Thr Trp Ser Pro Cys Ser Tyr Cys Ala Gly Lys
85 90 95

Leu Val Ala Leu Val Lys Ser Cys Pro His Leu Ser Leu Gln Ile Phe
100 105 110

Thr Ser Arg Leu Tyr Tyr His Trp Leu Trp Lys Asn Gln Ala Gly Leu
115 120 125

Arg Tyr Leu Trp Lys Ile Asn Ile Ser Val Leu Val Met Lys Glu Pro
130 135 140

Glu Phe Ala Asp Cys Trp Asp Asn Phe Val Asn His Gln Ser Arg Arg
145 150 155 160

```
Phe Lys Pro Trp Glu Lys Leu Thr Gln Tyr Ser Asn Ser Thr Glu Arg
                165                 170                 175

Arg Leu Leu Arg Ile Leu Arg Ile Asn Arg Thr Asp Leu Phe Leu Ala
            180                 185                 190

Gln Ser Ser Glu Gln Asp Pro Gly Leu Asn Asp Leu Val Asp Ala Ile
        195                 200                 205

Lys Arg Leu Phe Leu Asp Ala His Arg Pro Arg Asp
    210                 215                 220
```

<210> 104
<211> 198
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.37;
tr|A0A151P6M4|A0A151P6M4_ALLMI polypeptide sequence

<400> 104

```
Met Ala Val Glu Glu Glu Lys Gly Leu Leu Gly Thr Ser Gln Gly Trp
1               5                   10                  15

Lys Ile Glu Leu Lys Asp Phe Gln Glu Asn Tyr Met Pro Ser Thr Trp
            20                  25                  30

Pro Lys Val Thr His Leu Leu Tyr Glu Ile Arg Trp Gly Lys Gly Ser
        35                  40                  45

Lys Val Trp Arg Asn Trp Cys Ser Asn Thr Leu Thr Gln His Ala Glu
    50                  55                  60

Val Asn Cys Leu Glu Asn Ala Phe Gly Lys Leu Gln Phe Asn Pro Pro
65                  70                  75                  80

Val Pro Cys His Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Cys Gln
                85                  90                  95

Cys Cys Arg Arg Ile Leu Gln Phe Leu Arg Ala His Ser His Ile Thr
            100                 105                 110
```

```
Leu Val Ile Lys Ala Ala Gln Leu Phe Lys His Met Asp Glu Arg Asn
        115                 120                 125

Arg Gln Gly Leu Arg Asp Leu Val Gln Ser Gly Val His Val Gln Val
    130                 135                 140

Met Asp Leu Pro Asp Tyr Arg Tyr Cys Trp Arg Thr Phe Val Ser His
145                 150                 155                 160

Pro His Glu Gly Glu Gly Asp Phe Trp Pro Trp Phe Phe Pro Leu Trp
                165                 170                 175

Ile Thr Phe Tyr Thr Leu Glu Leu Gln His Ile Leu Leu Gln Gln His
            180                 185                 190

Ala Leu Ser Tyr Asn Leu
        195
```

<210> 105
<211> 158
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.38;
tr|A0A2K6MNR2|A0A2K6MNR2_RHIBE polypeptide sequence

<400> 105

```
Ile Trp Leu Cys Phe Thr Met Glu Ile Ile Lys Gln Cys Ser Thr Val
1               5                   10                  15

Ser Trp Lys Arg Gly Val Phe Arg Asn Gln Val Asp Pro Glu Thr His
            20                  25                  30

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Trp Glu Asp Thr Leu
        35                  40                  45

Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
    50                  55                  60

Cys Leu Asp Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
65                  70                  75                  80
```

```
Asn Val Lys Leu Ala Ile Phe Ala Ala Arg Leu Tyr Tyr Phe Trp Asp
                85                  90                  95


Thr Asp Tyr Gln Gln Gly Leu Arg Ser Leu Ser Glu Glu Gly Thr Ser
            100                 105                 110


Val Glu Ile Met Gly Tyr Glu Asp Phe Lys Tyr Cys Trp Glu Asn Phe
        115                 120                 125


Val Tyr Asn Gly Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
    130                 135                 140


Asn Phe Leu Phe Leu Asp Ser Lys Leu Gln Glu Ile Leu Glu
145                 150                 155


<210>  106
<211>  182
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.40;
       tr|F1CGT0|F1CGT0_ANOCA polypeptide sequence

<400>  106

Lys Ala Ala Ile Leu Leu Ser Asn Leu Phe Phe Arg Trp Gln Met Glu
1               5                   10                  15


Pro Glu Ala Phe Gln Arg Asn Phe Asp Pro Arg Glu Phe Pro Glu Cys
            20                  25                  30


Thr Leu Leu Leu Tyr Glu Ile His Trp Asp Asn Asn Thr Ser Arg Asn
        35                  40                  45


Trp Cys Thr Asn Lys Pro Gly Leu His Ala Glu Glu Asn Phe Leu Gln
    50                  55                  60


Ile Phe Asn Glu Lys Ile Asp Ile Lys Gln Asp Thr Pro Cys Ser Ile
65                  70                  75                  80


Thr Trp Phe Leu Ser Trp Ser Pro Cys Tyr Pro Cys Ser Gln Ala Ile
```

```
                85                  90                  95

Ile Lys Phe Leu Glu Ala His Pro Asn Val Ser Leu Glu Ile Lys Ala
            100                 105                 110

Ala Arg Leu Tyr Met His Gln Ile Asp Cys Asn Lys Glu Gly Leu Arg
        115                 120                 125

Asn Leu Gly Arg Asn Arg Val Ser Ile Met Asn Leu Pro Asp Tyr Arg
    130                 135                 140

His Cys Trp Thr Thr Phe Val Val Pro Arg Gly Ala Asn Glu Asp Tyr
145                 150                 155                 160

Trp Pro Gln Asp Phe Leu Pro Ala Ile Thr Asn Tyr Ser Arg Glu Leu
                165                 170                 175

Asp Ser Ile Leu Gln Asp
            180


<210>  107
<211>  230
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM3.41;
       tr|C7AGG3|C7AGG3_HORSE polypeptide sequence

<400>  107

Met Asp Pro Gln Ala Pro Thr Gln Arg Gly Gly Leu Gly Gln Ala Tyr
1               5                   10                  15

Gln Gly Gly Asp Tyr Val Gln Ala Pro Gly Asn Gly Asn Thr Gln His
            20                  25                  30

Leu Leu Ser Glu Asp Val Phe Lys Lys Gln Phe Gly Asn Gln Arg Arg
        35                  40                  45

Val Thr Lys Pro Tyr Tyr Arg Arg Lys Thr Tyr Val Cys Tyr Gln Leu
    50                  55                  60
```

```
Lys Leu Leu Arg Gly Pro Thr Ile Ala Lys Gly Tyr Phe Arg Asn Lys
65                  70                  75                  80

Lys Lys Arg His Ala Glu Ile Arg Phe Ile Asp Lys Ile Asn Ser Leu
                85                  90                  95

Gly Leu Asp Gln Asp Gln Ser Tyr Glu Ile Thr Cys Tyr Val Thr Trp
            100                 105                 110

Ser Pro Cys Ala Thr Cys Ala Cys Lys Leu Ile Lys Phe Thr Arg Lys
        115                 120                 125

Phe Pro Asn Leu Ser Leu Arg Ile Phe Val Ser Arg Leu Tyr Tyr His
    130                 135                 140

Trp Phe Arg Gln Asn Gln Gln Gly Leu Arg Gln Leu Trp Ala Ser Ser
145                 150                 155                 160

Ile Pro Val Val Val Met Gly Tyr Gln Glu Phe Ala Asp Cys Trp Glu
                165                 170                 175

Asn Phe Ala Asp Asn Arg Gly Asn Pro Phe Gln Ser Trp Glu Lys Leu
            180                 185                 190

Thr Glu Tyr Ser Lys Gly Ile Lys Arg Arg Leu Gln Lys Ile Leu Glu
        195                 200                 205

Pro Leu Asn Leu Asn Gly Leu Glu Asp Ala Met Gly Asn Leu Lys Leu
    210                 215                 220

Gly Ser Val Asp Leu Gly
225                 230
```

<210> 108
<211> 197
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.42;
tr|A0A250YMK7|A0A250YMK7_CASCN polypeptide sequence

<400> 108

```
Met Ser Leu Leu Lys Glu Asp Ile Phe Leu Tyr Gln Phe Asn Asn Gln
1               5                  10                  15

Gln Gln Val Gln Lys Pro Tyr Phe Arg Arg Arg Thr Tyr Leu Cys Tyr
            20                  25                  30

Gln Leu Glu Gln Pro Asn Gly Ser Arg Pro Gln Trp Pro Ala Lys Gly
        35                  40                  45

Cys Leu Gln Asn Lys Lys Gly His His Ala Glu Ile Arg Phe Ile Lys
    50                  55                  60

Arg Ile His Ser Met Gly Leu Glu Gln Asp Gln Asp Tyr Gln Ile Thr
65                  70                  75                  80

Cys Tyr Ile Thr Trp Ser Pro Cys Leu Ala Cys Ala Cys Ala Leu Ala
                85                  90                  95

Glu Leu Lys Asn His Phe Pro Arg Leu Thr Leu Arg Ile Phe Ala Ser
            100                 105                 110

Arg Leu Tyr Phe His Trp Ile Arg Lys Phe Gln Met Gly Leu Gln His
        115                 120                 125

Leu Tyr Lys Ser Gly Val Leu Val Ala Val Met Ser Leu Pro Glu Phe
    130                 135                 140

Thr Asp Cys Trp Glu Lys Phe Val Asn His Arg Gln Val Phe Phe Thr
145                 150                 155                 160

Pro Trp Asp Lys Leu Glu Glu His Ser Arg Ser Ile Gln Arg Arg Leu
                165                 170                 175

Arg Arg Ile Leu Gln Ser Trp Asp Val Asp Asp Leu Thr Asp Asp Phe
            180                 185                 190

Arg Asn Leu Arg Leu
        195
```

<210> 109

<211> 189
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM3.43;
tr|B7T160|B7T160_SHEEP polypeptide sequence

<400> 109

Met Pro Trp Ile Ser Asp His Val Ala Arg Leu Asp Pro Glu Thr Phe
1 5 10 15

Tyr Phe Gln Phe His Asn Leu Leu Tyr Ala Tyr Gly Arg Asn Cys Ser
20 25 30

Tyr Ile Cys Tyr Arg Val Lys Thr Trp Lys His Arg Ser Pro Val Ser
35 40 45

Phe Asp Trp Gly Val Phe His Asn Gln Val Tyr Ala Gly Thr His Cys
50 55 60

His Ser Glu Arg Arg Phe Leu Ser Trp Phe Cys Ala Lys Lys Leu Arg
65 70 75 80

Pro Asp Glu Cys Tyr His Ile Thr Trp Phe Met Ser Trp Ser Pro Cys
85 90 95

Met Lys Cys Ala Glu Leu Val Ala Gly Phe Leu Gly Met Tyr Gln Asn
100 105 110

Val Thr Leu Ser Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Lys Pro
115 120 125

Gln Tyr Arg Lys Gly Leu Leu Arg Leu Ser Asp Gln Gly Ala Cys Val
130 135 140

Asp Ile Met Ser Tyr Gln Glu Phe Lys Tyr Cys Trp Lys Lys Phe Val
145 150 155 160

Tyr Ser Gln Arg Arg Pro Phe Arg Pro Trp Lys Lys Leu Lys Arg Asn
165 170 175

Tyr Gln Leu Leu Ala Ala Glu Leu Glu Asp Ile Leu Gly
            180   185

<210> 110
<211> 541
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.1;
       tr|A0A182D0J1|A0A182D0J1_BLAVI polypeptide sequence

<400> 110

Met Thr Asn Pro Glu Ser Pro Pro Gln Ala Pro Cys Asp Phe Asn Glu
1   5   10   15

Asp Ala Leu Leu Asn Arg Glu Pro Leu Arg Gly Ser Pro Ile Lys Phe
            20   25   30

Val Ser Pro Val Asp Tyr Pro Asp Leu Val Phe Ala Leu Ala Gly Pro
        35   40   45

Val Gly Val Asp Ile Asp Tyr Ile Gln Gln Ser Ile Ser Asp Cys Leu
    50   55   60

Lys Ser Phe Asp Tyr Ser Thr Glu Phe Ile Arg Ile Thr Glu Ile Met
65   70   75   80

Gln Asp Ile Lys Cys Ser Lys Thr Ile Asp Cys Thr Asp Met Leu Lys
                85   90   95

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

Arg Ala Lys Asp Leu Leu Ala Ala Leu Thr Ile Ser Ala Ile Ser Lys
        115   120   125

Leu Arg Glu Gln Ile Lys Glu Arg Asp Glu Ala Thr Asn Lys Ser Asn
    130   135   140

Ile Gln Pro Ser Arg Arg Lys Leu Ala Trp Val Val Arg Gln Leu Lys
145   150   155   160

```
Thr Pro Glu Glu Val Arg Leu Leu Arg Ala Val Tyr Gly Lys Gln Phe
                165                 170                 175

Val Leu Val Ser Ile Tyr Ser Ser Pro Gln Arg Arg Glu Asp Phe Leu
            180                 185                 190

Ile Ser Lys Ile Lys Ile Lys Ser Arg Gly Thr Ile Asp Asn Asn Thr
        195                 200                 205

Ser Ser Glu Gly Ala Gln Arg Leu Ile Glu Arg Asp Ser Lys Glu Asp
    210                 215                 220

Asn Glu Tyr Gly Gln Asn Leu Ser Gly Thr Phe Cys Leu Gly Asp Ile
225                 230                 235                 240

Phe Val Asp Ser Asn Asn Lys Glu Ser Ala Ile Val Ser Ile Asp Arg
                245                 250                 255

Phe Leu Asn Ala Phe Phe Gly Ser Asn Glu Ile Ser Pro Thr Arg Asp
            260                 265                 270

Glu Tyr Gly Met Tyr Leu Ala Lys Thr Ala Ser Leu Arg Ser Cys Asp
        275                 280                 285

Leu Ser Arg Gln Val Gly Ala Ala Ile Phe Ser Lys Thr Gly Glu Ile
    290                 295                 300

Ile Ser Leu Gly Ser Asn Glu Val Pro Lys Ala Gly Gly Gly Thr Tyr
305                 310                 315                 320

Trp Thr Gly Asp Asn Ala Asp Ser Arg Asp Ile Arg Leu Gly His Asp
                325                 330                 335

Pro Asn Glu Ile Asn Lys Val Glu Ile Phe Ala Glu Ile Ile Ser Arg
            340                 345                 350

Leu Leu Glu Asp Lys Leu Leu Ser Asn Asp Leu Leu Asn Lys Asp Ala
        355                 360                 365

Ala Ser Ile Val Thr Ile Leu Leu Ser Lys Asn Glu Gly Lys Arg Tyr
```

```
    370                 375                 380

Lys Asp Leu Arg Val Met Asp Ile Ile Glu Phe Gly Arg Ile Ile His
385                 390                 395                 400

Ala Glu Met Ser Ala Ile Cys Asp Ala Ala Arg Asn Gly Arg Ala Ile
                405                 410                 415

Ile Gly Ala Thr Leu Phe Cys Thr Thr Phe Pro Cys His Leu Cys Ala
            420                 425                 430

Lys His Ile Val Ala Ser Gly Ile Gly Arg Ile Val Tyr Leu Glu Pro
        435                 440                 445

Tyr Pro Lys Ser Tyr Ala Lys Lys Leu His Ser Asp Ser Ile Gln Val
    450                 455                 460

Glu Asp His Ser Asp Ser Glu Lys Val Ser Phe Glu Pro Phe Ile Gly
465                 470                 475                 480

Ile Ser Pro Ser Arg Tyr Arg Glu Leu Phe Glu Gly Gly Arg Arg Lys
                485                 490                 495

Asp Pro Phe Gly Glu Ala Leu Lys Trp Lys Asn Asp Pro Arg Lys Pro
            500                 505                 510

Val Ile Asp Val Val Val Pro Pro His Phe Glu Ala Glu Lys Leu Val
        515                 520                 525

Ile Ala Gln Leu Gly Lys Leu Ile Val Ser Gly Thr Gly
    530                 535                 540
```

<210> 111
<211> 330
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.2; tr|A0A2D6EXD2|A0A2D6EXD2_9ARCH polypeptide sequence

<400> 111

```
Met Ile Ile Gly Leu Val Gly Thr Ile Gly Ala Gly Lys Gln Thr Ile
1               5                  10                  15

Ile Asp Tyr Leu Gln Glu Lys Tyr Gly Tyr Asn Ala Leu Ser Cys Ser
            20                  25                  30

Asp Val Leu Arg Glu Ile Leu Lys Lys Gln Gly Lys Pro Val Thr Arg
        35                  40                  45

Asp Asn Leu Arg Glu Ile Gly Asn Lys Thr Arg Glu Glu Gly Gly Asn
    50                  55                  60

Gly Ala Ile Ala Lys Ile Leu Leu Glu Lys Leu Arg Asn Asn Trp Lys
65                  70                  75                  80

Ala Asn Tyr Ile Val Asp Ser Leu Arg His Pro Asp Glu Val Ser Val
                85                  90                  95

Leu Arg Thr Ser Pro Leu Phe His Leu Val Ala Val Asp Ala Asp Leu
            100                 105                 110

Arg Ile Arg Phe Glu Arg Val Lys Ala Arg Lys Arg Glu Glu Glu Pro
        115                 120                 125

Thr Thr Leu Pro Ala Phe Val Glu Arg Asp Gln Lys Glu Met Phe Gly
    130                 135                 140

Thr Gly Asn Glu Gln Arg Ile Arg Glu Thr Met Glu Leu Ala Asp Glu
145                 150                 155                 160

Leu Val Leu Asn Asn Gly Thr Val Glu Glu Leu Lys Gln Arg Ile Asp
                165                 170                 175

Asp Leu Asn Leu Val Ser Asp Glu Arg Leu Arg Pro Ser Trp Asp Asp
            180                 185                 190

Tyr Phe Met Arg Leu Ala Arg Leu Ala Ala Gln Arg Ser Asn Cys Met
        195                 200                 205

Ser Arg Lys Val Gly Ala Ile Ile Thr Lys Asp Arg Arg Val Ile Ala
    210                 215                 220
```

```
Thr Gly Tyr Asn Gly Thr Pro Arg Gly Val Lys Asn Cys Asn Glu Gly
225                 230                 235                 240

Gly Cys Glu Arg Cys Asn Ser Ala Val Ala Lys Gly Thr Ala Ile Ser
                245                 250                 255

Glu Cys Leu Cys Leu His Gly Glu Glu Asn Ala Ile Ile Glu Ala Gly
            260                 265                 270

Arg Val Arg Ser Glu Gly Ala Thr Ile Tyr Thr Ser Phe Leu Pro Cys
        275                 280                 285

Leu Trp Cys Thr Lys Met Ile Ile Gln Ala Gly Leu Lys Glu Val Val
    290                 295                 300

Phe Ser Glu Val Tyr Asp Leu His Glu Ala Ser Ile Lys Leu Phe Glu
305                 310                 315                 320

Thr Ser Gly Val Leu Ile Arg Arg Leu Lys
                325                 330
```

<210> 112
<211> 367
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.3;
tr|F7YVM7|F7YVM7_9THEM polypeptide sequence

<400> 112

```
Met Asn Glu Phe Lys Tyr Met Ser Leu Ala Leu Lys Leu Ala Lys Lys
1               5                   10                  15

Gly Lys Tyr Thr Thr Ser Pro Asn Pro Met Val Gly Ala Val Ile Val
            20                  25                  30

Lys Asp Gly Lys Ile Leu Ala Thr Gly Tyr His Lys Lys Ala Gly Gln
        35                  40                  45

Pro His Ala Glu Ile Asn Ala Leu Ser Lys Leu Asn Phe Gln Ala Gln
```

```
    50                  55                  60

Asn Cys Glu Met Tyr Val Thr Leu Glu Pro Cys Ser His Tyr Gly Arg
65                  70                  75                  80

Thr Pro Pro Cys Ala Asp Ala Ile Ile Arg Ser Gly Ile Arg Lys Val
                85                  90                  95

Val Ile Ala Thr Leu Asp Pro Asn Pro Leu Val Asn Gly Lys Gly Val
            100                 105                 110

Glu Lys Leu Lys Asn Ala Gly Ile Glu Val Val Cys Gly Val Leu Glu
        115                 120                 125

Glu Lys Ala Lys Lys Leu Asn Glu Lys Phe Phe Lys Tyr Ile Thr Thr
    130                 135                 140

Lys Ile Pro Phe Val Ala Leu Lys Ile Ala Gln Thr Leu Asp Gly Lys
145                 150                 155                 160

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

Arg Glu Tyr Val His Lys Leu Arg Met Glu Tyr Asp Ala Val Leu Thr
            180                 185                 190

Gly Ile Gly Thr Ile Leu Lys Asp Asp Pro Gln Leu Asn Val Arg Leu
        195                 200                 205

Lys Lys Val Tyr Lys Gln Pro Leu Arg Ile Ile Leu Asp Ser Lys Leu
    210                 215                 220

Lys Ile Pro Leu Ser Ala Lys Val Leu Glu Asp Pro Ser Lys Val Ile
225                 230                 235                 240

Ile Leu Thr Thr Ala Leu Ala Asp Lys Glu Lys Leu Glu Glu Leu Arg
                245                 250                 255

Ser Lys Gly Val Glu Val Ile Ile Thr Asn Glu Lys Asn Gly Ile Val
            260                 265                 270
```

```
Asp Leu Glu Ser Ala Leu Lys Ile Leu Gly Glu Lys Lys Ile Thr Ser
        275                 280                 285

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

Ser Leu Phe Asp Lys Ile Tyr Leu Phe Ile Ala Pro Lys Ile Phe Gly
305                 310                 315                 320

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

Lys Ser Gln Lys Phe Ser Leu Glu Ser Val Lys Lys Ile Gly Glu Asp
            340                 345                 350

Leu Leu Leu Glu Leu Tyr Pro Lys Gln Leu Lys Lys Leu Glu Glu
        355                 360                 365
```

<210> 113
<211> 449
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.4; tr|A0A3M6UNF1|A0A3M6UNF1_9CNID polypeptide sequence

<400> 113

```
Met Glu Glu Lys Ser Glu Leu Glu Asn Glu Leu Met Arg Ser Thr Ser
1               5                   10                  15

Pro Lys Pro Ser Val Pro Asn Gly Ser Lys Gly Asn Glu Cys Glu Gln
            20                  25                  30

Arg Glu Thr Arg Ile Thr Lys Glu Asn Leu Tyr Met Val Leu Ala Leu
        35                  40                  45

Trp Met Glu Glu Phe Pro Val Val Glu Gln Thr Ser Ser Ala Lys Arg
    50                  55                  60

Leu Asn Lys Val Gly Val Val Phe Val Leu Pro Thr Asp Arg Val Leu
65                  70                  75                  80
```

```
Ala Ala Asp Cys Ser Arg Asp Gly Val His Gly Val Ala Arg Val Met
               85                  90                  95

Val Asn His Cys Gly Lys Leu Glu Gly Cys Lys Val Phe Val Ser Arg
            100                 105                 110

Lys Pro Cys Ser Leu Cys Ala Lys Leu Leu Val Gln Ser Lys Val Ser
        115                 120                 125

Arg Val Phe Tyr Leu Pro Ile Glu Pro Glu Ser Glu Asn Lys Gly Glu
    130                 135                 140

Ile Ala Arg Ala Asp Asn Leu Phe Lys Asn Ser Ser Val Gly Gln Ser
145                 150                 155                 160

Val Phe Val Pro Cys Val Glu Gln Lys Val Leu Asp Lys Leu Glu Asp
                165                 170                 175

Lys Leu Pro Lys Glu Ile Ile Thr Pro Asp Asp Ile Ser Glu Cys Arg
            180                 185                 190

Asp Asn Leu Leu Lys Lys Cys Gly Trp Ser Ala Glu Trp Phe Ala Arg
        195                 200                 205

Ala Gln Ala Ser Leu Pro Trp Pro Cys Phe Glu Gly Lys Met Lys Ser
    210                 215                 220

Gln Val Asp Asn Asp Phe Lys Ser Leu Ile Lys Trp Ile Ala Val Val
225                 230                 235                 240

Lys Ala Pro Met Asp Lys Gly Val Ala Phe Pro Lys Val Lys Leu Thr
                245                 250                 255

Ser Asp Ser Arg Val Val Pro Asp Cys Asp Ala Asp Asn Phe Pro Asp
            260                 265                 270

Ser Lys Thr Ala Tyr His Met Met Ile Phe Ala Lys Met Leu Ala Arg
        275                 280                 285
```

```
Gln Thr Asp Asp Pro Lys Thr Gly Val Gly Ala Val Ile Val Arg Gly
    290                 295                 300

Lys Val Pro Asp Ile Val Ser Leu Gly Trp Asn Gly Phe Pro Ser Lys
305                 310                 315                 320

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

Gln Lys Lys Phe Pro Tyr Val Ile His Ala Glu Gln Asn Ala Leu Met
            340                 345                 350

Val Arg Asn Val Lys Asp Leu Thr Asp Gly Ile Leu Phe Val Thr Lys
        355                 360                 365

Pro Pro Cys Asp Glu Cys Ala Pro Met Ile Lys Leu Ser Gly Val Lys
    370                 375                 380

Thr Ile Val Ile Gly Glu Lys Ile Glu Lys Ser Arg Gly Gly Glu Leu
385                 390                 395                 400

Ser Tyr Asn Leu Ile Lys Glu Tyr Ile Lys Glu Gly Ile Met Thr Cys
                405                 410                 415

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

Pro Glu Thr Arg Lys Arg Leu Lys Ser Ser Cys Ser Asn Ser Asn Asp
        435                 440                 445

Val
```

<210> 114
<211> 414
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.5; tr|A0A2G3K826|A0A2G3K826_9BURK polypeptide sequence

<400> 114

```
Met Thr Lys Ile Ile Asp Asp Val Asn Thr Ala Ala Ala Ala Val Leu
1               5                  10                  15

Asp Gln Ala Thr Ala Ala Ala Asn Gln Thr Thr Phe Ala Val Gly Gly
            20                  25                  30

Val Met Val Asn Asn Gln Thr Gly Glu Val Ile Ser Ala Ile His Asn
        35                  40                  45

Asn Val Ile Ile Pro Leu Ser Asn Asn Val Ser Phe Thr Phe Asp Pro
    50                  55                  60

Thr Ala His Gly Glu Arg Gln Leu Val Tyr Trp Tyr Tyr Ala Asn Lys
65                  70                  75                  80

Glu Ala Leu Lys Leu Pro Glu Pro Asn Gln Ile Thr Val Ile Thr Ser
                85                  90                  95

Leu Asp Pro Cys Ala Met Cys Thr Gly Ala Leu Leu Thr Ala Gly Phe
            100                 105                 110

Asn Val Gly Val Val Ala Ile Asp Thr Tyr Ala Gly Ile Asn Cys Ala
        115                 120                 125

Gln Asn Phe Gln Phe Ala Thr Leu Pro Ala Asn Leu Arg Thr Lys Ala
    130                 135                 140

Gln Lys Asn Phe Gly Tyr Tyr Ala Ser Gly Ala Ala Asn Phe Lys Pro
145                 150                 155                 160

Leu Thr Arg Ser Tyr Val Gly Gly Pro Ser Val Ala Phe Lys Asn Gly
                165                 170                 175

Val Val Thr Pro Ala Asn Leu Arg Asp Cys Gly Thr Val Phe Thr Gln
            180                 185                 190

Ser Val Asp Thr Val Arg Asn Thr Ser Asn Ser Thr Gly Leu Ala Pro
        195                 200                 205

Ser Gln Met Ser Asn Pro Ala Glu Leu Pro Ser Asn Ser Ala Ile Leu
```

```
    210                 215                 220

Gln Ala Tyr Arg Ala Ile Tyr Lys Lys Ala Phe Thr Ile Lys Ile Asp
225                 230                 235                 240

Asn Pro Arg Leu Pro Asp Ala Gln Ile Leu Thr Glu Leu Lys Ala Val
                245                 250                 255

Leu Ala Asp Ala Pro Asn Ala Arg Asn Ala Val Ala Phe Ile Asp Pro
            260                 265                 270

Phe Gly Asn Leu Val Leu Cys Met Ala Asp Ala Phe Asn Thr Ser Pro
        275                 280                 285

Val His Ala Ala Phe Met Asn Val Thr Gln Glu Tyr Ala Lys Thr Arg
    290                 295                 300

Trp Asp Leu Met Asn Lys Tyr Ala Gln Ala Ser Thr Thr Asp Asn Pro
305                 310                 315                 320

Ala Leu Tyr Leu Thr His Pro Lys Tyr Gly Thr Phe Val Tyr Leu Tyr
                325                 330                 335

Ala Pro Asp Pro Asp Asp Ser Ile Thr Ile Met Ser Leu Gly Ala Tyr
            340                 345                 350

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

Gln Phe Tyr Tyr Pro Pro Arg Asn Gly Ala Gln Phe Ser Glu Leu Val
    370                 375                 380

Pro Val Val Asn Glu Leu Pro Pro Phe Tyr Thr Gln Asn Val Asn Ile
385                 390                 395                 400

Ser Leu Met Gln Val Pro Gly Val Thr Gln Ala Pro Thr Lys
                405                 410

<210>  115
<211>  308
<212>  PRT
```

<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.6;
tr|K1ZCJ4|K1ZCJ4_9BACT polypeptide sequence

<400> 115

Met Ser Ser Arg Ala Lys Lys Asn Arg Ser Thr Asn Leu Lys Lys Ser
1 5 10 15

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

Gln Val Leu Val Ala Tyr Val Pro Val Ile His Glu Gly Tyr Arg Arg
35 40 45

Phe Phe Arg His Phe Pro Ala Val Lys Glu Leu Trp Leu Ile Ser Gln
50 55 60

Glu Leu Ser His Glu Leu Arg Ser Leu Gln Lys Asp Ile Arg Ala Leu
65 70 75 80

Lys Ala Ser Glu Thr Lys Lys Leu Leu Gln Thr Trp Gly Gln Phe Gln
85 90 95

Lys Ile Lys Leu Leu Thr Pro Ser Ser Leu Ala Ile Leu Gln Lys Thr
100 105 110

Thr Thr Gln Leu Val Phe Pro Asp Glu Glu Ile Ser His His Leu Val
115 120 125

Glu Lys Tyr Phe Ala Gln Asn Arg Val Leu Phe Ala Ser Phe Phe Leu
130 135 140

Arg Trp Asp Lys Lys Ser Ser Leu Lys Lys His Asp Leu Gln Glu Tyr
145 150 155 160

Ser Glu Ile Ser Asn Lys Glu Phe Asp Gln Met Met Ile Ala Ile Ala
165 170 175

Gln Gln Glu Ala Asp Lys Ser Asp Asp Trp Trp Arg Gln Val Gly Gly
180 185 190

```
Leu Ile Phe Lys Asp Glu Thr Ile Leu Leu Leu Ala His Asn Gln His
        195                 200                 205


Thr Pro Thr Glu Ala Glu Ala Tyr Phe Ala Gly Asp Pro Arg Ala Asp
    210                 215                 220


Phe His Gln Gly Glu Tyr Leu Lys Ile Ser Thr Ala Ile His Ala Glu
225                 230                 235                 240


Ala Tyr Leu Ile Ala Gln Ala Ala Lys Gln Gly Ile Ser Leu Glu Gly
                245                 250                 255


Ala Asp Leu Tyr Val Thr Thr Phe Pro Cys Pro Val Cys Ala Lys Gln
            260                 265                 270


Val Ala Tyr Ser Gly Ile Lys Arg Val Phe Phe Arg Glu Gly Tyr Ser
        275                 280                 285


Leu Leu Asp Gly Glu Thr Ile Leu Lys Ala Asn Gly Val Lys Leu Ile
    290                 295                 300


Arg Val Thr Val
305
```

<210> 116
<211> 615
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.7;
       tr|A0A1G3PNQ8|A0A1G3PNQ8_9SPIR polypeptide sequence

<400> 116

```
Met Arg Asp Leu Pro Leu Leu Val Leu Gly Leu Thr Gly Pro Met Gly
1               5                   10                  15


Ala Gly Cys Thr Arg Phe Ala Arg Asp Ile Ser Lys Met Glu Pro Gly
            20                  25                  30


Lys Val Ile Lys Lys Gln Gly Leu Leu Asp Gln Val Ala His Glu Ile
```

```
        35                  40                  45

Ser Glu Leu Ser Lys Lys Ala Ser Glu Ile Arg Leu Gln Cys Ile Ser
    50                  55                  60

Asn Gly Lys Asn Ser Glu Leu Ala Glu Leu Lys Arg Leu Asn Arg Arg
65                  70                  75                  80

Leu Asn Ala Lys Leu Ala Glu Arg Ala Cys Leu His Val Ile Ala Lys
                85                  90                  95

Ser Ser Leu Pro Glu Pro Leu Phe Ile Ser Leu Asn Thr Ile Val Ile
            100                 105                 110

Lys Ile Ala Val Asp Ser Ile Thr Ala Pro Glu Phe Ala Glu Trp Ala
        115                 120                 125

Lys Asn His Ala Lys Val Ala Asp Leu Leu Lys Trp Leu Arg Thr Gln
    130                 135                 140

Trp Glu Ser Glu Leu Thr Leu Tyr Glu Thr Trp Gly Gln Asp Ala Gly
145                 150                 155                 160

Arg Phe Ser Gln Asp Glu Leu Glu Lys Met Asp Ala Met Phe Ala Glu
                165                 170                 175

Phe Glu Arg Ile Gly Asp Glu Ile Leu Lys Glu Asp Phe Glu Thr Tyr
            180                 185                 190

Phe Gly Lys Arg Asn Asn Asp Phe Ser Ile Arg Met Phe Ser Glu Asn
        195                 200                 205

Ile Arg Leu Ser Gly Asn Pro Phe Arg Pro Ala Glu Asn Gly Gly Gly
    210                 215                 220

Gly Gly Lys Tyr Asp Glu Pro Ser Met Val Met Ile Ala Arg Glu Thr
225                 230                 235                 240

Asp Arg Tyr Ile Arg Phe Tyr Arg Thr Arg Ser Asp Gln Lys Arg Ser
                245                 250                 255
```

```
His Phe Phe Ile Ile Asp Glu Ile Lys Asn Pro Arg Glu Ala Glu Tyr
            260                 265                 270

Phe Arg Ala Arg His Gln Asn Phe Phe Leu Val Ser Ile Phe Ser Ser
        275                 280                 285

Ser Glu Ile Arg Ala Ser Arg Met Arg Arg Gly Leu Gly His Asp Ala
    290                 295                 300

Gly Val Ser Asp Ala Asp Phe Gln His Leu Phe Arg Glu Leu Asp Ser
305                 310                 315                 320

Arg Asp Trp Gly Ala Asp Asp Phe Asp Ala His Gly Leu His Arg Gln
                325                 330                 335

Asn Ile Tyr Arg Cys Phe Asn Leu Ala Asp Ile Ala Ile Asn Asn Asp
            340                 345                 350

Val Glu Asp Glu Arg Phe Ser Glu Val Leu Phe Asn Lys Phe Ile Arg
        355                 360                 365

Tyr Tyr Ala Leu Met Leu Ser Pro Gly Cys Val Gln Pro Thr Pro Gln
    370                 375                 380

Glu Thr Tyr Met His Leu Ala Tyr Ser Leu Ser Leu Arg Ser Thr Cys
385                 390                 395                 400

Ile Ser Arg Gln Val Gly Ala Val Ile Thr Asp Leu Glu Asp Arg Ile
                405                 410                 415

Leu Ser Leu Gly Trp Asn Glu Val Pro Glu Gly Gln Ile Gly Cys Gly
            420                 425                 430

Leu Lys Val Lys Lys Asp Tyr Thr Asp Lys Glu Asn Pro Leu Phe Glu
        435                 440                 445

Met Glu Ile Trp Asp Asn Val Ile Thr Ala Glu Asp Leu Ala Val Trp
    450                 455                 460

Asp Asp Glu Asp Ser Ile Cys Val Lys Asp Ile Leu Ser Arg Ile Glu
```

```
465                 470                 475                 480

Ile Lys Thr Lys Leu Lys Ser Val Ser Leu Thr Pro Glu Glu Arg Ala
                485                 490                 495

Asp Val Leu Lys Ala Leu Arg Ile Lys Arg Leu Glu Tyr Ser Arg Ser
            500                 505                 510

Leu His Ala Glu Glu Asn Ala Ile Leu Gln Val Ala Ser Arg Gly Gly
        515                 520                 525

Val Gly Leu Lys Asp Gly Thr Ile Tyr Val Thr Thr Phe Pro Cys Glu
    530                 535                 540

Leu Cys Ser Lys Lys Ile Tyr Gln Val Gly Ile Ser Lys Ile Tyr Tyr
545                 550                 555                 560

Thr Glu Pro Tyr Pro Asn Ser Ile Ser Glu Lys Val Ile Leu Lys Asp
                565                 570                 575

Gly Ile Arg Asn Ile Lys Ile Leu Gln Phe Glu Gly Val Lys Ser Tyr
            580                 585                 590

Ser Tyr Phe Lys Leu Phe Lys Pro Gly Phe Asp Lys Lys Asp Ala Gln
        595                 600                 605

Met Leu Glu Gly Arg Gly Ile
    610                 615
```

<210> 117
<211> 291
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.8; tr|A0A1G0PGF4|A0A1G0PGF4_9BACT polypeptide sequence

<400> 117

```
Met Lys His Asn Asn Gln Leu Arg Lys Glu Ile Glu Lys Leu Leu Gly
1               5                   10                  15
```

```
Gln Asn Ser Ile Ile Lys Asn Asp Glu Leu Lys Lys Leu Gln Lys Glu
            20                  25                  30

Tyr Lys Ile Glu Thr Asp Glu Leu Leu Ile Ser Phe Leu Pro Tyr Ala
        35                  40                  45

Ala Glu Phe Ala Lys Val Pro Ile Ser Lys Tyr Lys Val Gly Ala Val
    50                  55                  60

Val Leu Gly Lys Ser Gly Asn Ile Tyr Phe Gly Ser Asn Met Glu Phe
65                  70                  75                  80

Glu Ala Gly Ala Leu Ser Ala Thr Val His Ala Glu Gln Ser Ala Val
                85                  90                  95

Asn Asn Ala Trp Leu Asn Gly Glu Thr Gly Ile Asn Lys Ile Ala Val
            100                 105                 110

Thr Ala Ala Pro Cys Gly Tyr Cys Arg Gln Phe Leu Asn Glu Leu Thr
        115                 120                 125

Thr Ala Lys Gln Leu His Val Leu Leu Lys Asp Lys Asn Leu Glu Ala
    130                 135                 140

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

Arg Asp Leu Glu Ile Glu Gly Gly Leu Met Lys Val Glu Asn His Lys
                165                 170                 175

Leu Lys Ile Glu Asn Ile Asn Asp Glu Leu Ile Asn Ala Ala Leu Glu
            180                 185                 190

Ala Ala Asn Lys Ser Tyr Ala Pro Tyr Ser Lys Asn Tyr Ser Gly Val
        195                 200                 205

Ser Ile Gln Leu Ser Asp Gly Thr Ile Phe Ser Gly Arg Tyr Ser Glu
    210                 215                 220

Asn Ala Ala Tyr Asn Pro Ser Leu Leu Pro Phe Gln Ser Ala Leu Ala
225                 230                 235                 240
```

Phe Met Asn Met Asn Thr Lys Lys Gly Ser Asn Asn Lys Ile Val Asp
245 250 255

Ala Val Leu Val Glu Ala Val Ser Asn Ile Ser Gln Lys Asp Ala Ala
260 265 270

Gly Thr Leu Leu Asn Ser Ile Ser Lys Thr Lys Leu Arg Tyr Tyr Lys
275 280 285

Ile Lys Asn
290

<210> 118
<211> 398
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.9;
tr|A0A0P4WGY5|A0A0P4WGY5_9EUCA polypeptide sequence

<400> 118

Met Glu Glu Asn Ser Ser Ala Thr Ser Gln Pro Lys Cys Ala Ser Arg
1 5 10 15

Thr Lys Gln Gly Gly Asn Asp Leu Ser Thr Asp Met Ser Asn Leu Ser
20 25 30

Val Gly Glu Thr Lys Arg Thr Asp Phe Leu Pro Trp Asp Asp Tyr Phe
35 40 45

Met Ala Val Ala Phe Leu Ser Ala Met Arg Ser Lys Asp Pro Ser Ser
50 55 60

Gln Val Gly Ala Cys Ile Val Asn Ala Asp Lys Lys Ile Val Gly Ile
65 70 75 80

Gly Tyr Asn Gly Met Pro Ile Gly Cys Ser Asp Asp Glu Leu Pro Trp
85 90 95

Asn Lys Glu Ser Leu Asp Pro Leu Gln Thr Lys Tyr Met Tyr Val Cys

```
            100                 105                 110

His Ala Glu Met Asn Ala Ile Met Asn Lys Asn Ser Ser Asp Leu Ala
        115                 120                 125

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

Leu Val Ile Gln Ala Gly Ile Arg Glu Val Val Phe Phe Ser Asp Lys
145                 150                 155                 160

His Gln Gln Lys Pro Glu Thr Val Ala Ser Lys Lys Met Leu Asn Met
                165                 170                 175

Ala Gly Val Ala Tyr Arg Gln Tyr Thr Pro Ser Gln Ser Lys Ile Glu
            180                 185                 190

Leu Asn Leu Ser Leu Lys Glu Gln Glu Lys Ser Glu Pro Thr Ala Asp
        195                 200                 205

Ile Thr Gln Ser Ser Glu Arg Asp Gln Asn Ser Lys Arg Lys Asp Tyr
    210                 215                 220

Leu Ser Trp Glu Glu Tyr Phe Met Ala Met Ala His Leu Ser Ala Leu
225                 230                 235                 240

Arg Ser Lys Asp Pro Ile Thr Gln Val Gly Ala Cys Ile Val Asn Ser
                245                 250                 255

Lys Lys Lys Ile Val Gly Ile Gly Tyr Asn Gly Met Pro Leu Gly Cys
            260                 265                 270

Asn Asp Asp Leu Met Pro Trp Gly Asn Ser Ser Ser Asn Lys Leu Glu
        275                 280                 285

Thr Lys Tyr Met Tyr Val Cys His Ala Gly Val Asn Ala Ile Met Asn
    290                 295                 300

Lys Asn Ser Cys Asp Val Ser Gly Cys Thr Leu Tyr Val Ala Leu Phe
305                 310                 315                 320
```

```
Pro Cys Asn Glu Cys Ala Lys Val Ile Ile Gln Ala Gly Ile Lys Thr
                325                 330                 335

Ile Ile Tyr Ala Ser Asp Thr Asn Lys Asp Gln Ala Ser Ile Leu Ala
            340                 345                 350

Ser Lys Lys Met Leu Asp Met Ala Gly Ile Lys Tyr Arg Ala Asp Asn
        355                 360                 365

Leu Ser Gln Arg Lys Ile Val Ile Asp Phe Lys Thr Ile Asp Trp Asn
    370                 375                 380

Ser Arg Phe Met Asn Asp His Gln Asn Asp Pro Thr Cys Leu
385                 390                 395
```

<210> 119
<211> 259
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.10; tr|A0A3D8IG27|A0A3D8IG27_9HELI polypeptide sequence

<400> 119

```
Met Arg Lys Asn Ile Leu Tyr Phe Ile Leu Thr Leu Phe Phe Leu Ser
1               5                   10                  15

Gly Leu Tyr Ala Thr Ser Leu Pro Glu Asp Asn Val Val Ser Gly Val
            20                  25                  30

Ile Tyr Glu Lys Ile Asp Thr Val Ser Ala Glu Val Asp His Ile Tyr
        35                  40                  45

Pro Met Leu Ala Leu Ala Ile Val Tyr Lys Asp Trp Gln Glu Lys Asn
    50                  55                  60

Met Leu Asn Lys Gln Gly His Asn Ile Gly Leu Val Ile Val Asp Glu
65                  70                  75                  80

Asn Asn Met Pro Val Phe Trp Val Arg Asn Ser Val His Ala Thr His
                85                  90                  95
```

```
Asn Gly Thr Gln His Gly Glu Val Arg Leu Val Ser Asn Leu Leu Asn
            100                 105                 110

Cys Glu Gly Phe Asn Lys Tyr Leu Asp Lys Tyr Thr Leu Tyr Thr Thr
        115                 120                 125

Leu Glu Pro Cys Ile Met Cys Ala Gly Met Leu Ser Met Val Gln Ile
    130                 135                 140

Pro Lys Val Val Tyr Ala Gln Lys Asp Leu Ser Cys Gly Asn Thr Gln
145                 150                 155                 160

Glu Ile Ile Ser Thr Ala Lys Tyr Pro Arg Tyr Tyr Lys Ala Phe Thr
                165                 170                 175

Val Glu Asn Gly Tyr Lys Lys Asp Leu Glu Glu Cys Phe Glu Gln Tyr
            180                 185                 190

Lys Ile Cys Lys Asn Asp Ser Ile Thr Asp Phe Leu Val Asn Asp Ser
        195                 200                 205

Ala Lys Glu Ile Phe Arg Lys Ala Ser Asn Asp Leu Gln Asp Tyr Lys
    210                 215                 220

Val Lys Phe Lys Glu Asn Arg Arg Val Ile Lys Val Ala Gln Glu Phe
225                 230                 235                 240

Leu Gln Asn Ile Gln Thr Lys Asp Asn Leu Asp Val Leu Gln Cys Pro
                245                 250                 255

Lys Asn Met
```

<210> 120
<211> 256
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.11;
tr|A0A351C8C4|A0A351C8C4_9BACT polypeptide sequence

<400> 120

```
Met Asn Glu Leu Thr Lys Gln Ser Glu His Leu Arg Asn Glu Ala Leu
1               5                  10                  15

Arg Ile Ala Thr Arg Ser Tyr Val Pro Tyr Thr Gly Gln Gln Glu Gly
            20                  25                  30

Val Ile Ile Leu Leu Glu Asn Gly Asp Leu Ile Pro Gly Val Arg Val
        35                  40                  45

Glu Asn Ala Ser Phe Gln Leu Thr Ile Pro Ala Leu Gln Asn Ala Leu
    50                  55                  60

Ser Thr Met Tyr Ala Leu Gln Arg Thr Asp Ile Ser Met Ile Val Ser
65                  70                  75                  80

Ser Ile Pro Phe Thr Asp Ser Asp Leu Ala Tyr Thr Gly Gly Met Ala
                85                  90                  95

Glu Ile Ala Trp Glu Met Val Gly Ala Ser Leu Leu Leu Val Ala Gly
            100                 105                 110

Ala His Ile Pro Glu Ala Gly Thr Phe Ile Asp Pro Ala Arg Gly Glu
        115                 120                 125

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

Glu Ser Asp Phe Pro Val Gly Ser Ala Ile Gln Thr Ser Asp Asp Val
145                 150                 155                 160

Val Ile Asp Gly Cys Asn Val Glu His Ser Asp Trp Ser Lys Ile Ile
                165                 170                 175

Cys Ala Glu Arg Asn Val Leu Ser Thr Ala Arg Ser Tyr Gly Leu Gly
            180                 185                 190

Gln Ile Thr Thr Ile Tyr Val Ser Cys Pro Lys Glu Pro Gly Gly Thr
        195                 200                 205
```

Pro Cys Gly Ala Cys Arg Gln Val Ile Val Glu Leu Ala Pro Asp Ala
210 215 220

Thr Val Trp Met Asp Arg Gly Asn Gln Glu Pro Ile Ala Met Lys Ala
225 230 235 240

Thr Lys Leu Leu Pro Gly His Phe Thr Gly Asn Val Leu Lys Lys Gln
245 250 255

<210> 121
<211> 301
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.12;
tr|A0A1G6V2K7|A0A1G6V2K7_PEPNI polypeptide sequence

<400> 121

Met Pro Ile Val Arg Val Asn Glu Ile Gly Ala Arg Leu Pro Glu Asp
1 5 10 15

Trp Glu Ala Leu Glu Thr Ala Ile Trp Gln Ala Tyr Val Ser Arg Glu
20 25 30

Asp Leu Pro Asp Ala Gly Glu Leu Asp Leu Thr Leu Val Asp Asp Ala
35 40 45

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

Asp Val Leu Ser Phe Pro Met Tyr Asp Asp Arg Asp Asp Leu Ala Ala
65 70 75 80

Asp Val Gln Ala Gly Leu Pro Val Ile Leu Gly Asp Ile Met Ile Ser
85 90 95

Val Pro Thr Ala Glu Arg Gln Ala Gln Ala Tyr Gly His Ser Phe Lys
100 105 110

Arg Glu Met Ala Tyr Leu Leu Val His Gly Leu Leu His Ile Ala Gly
115 120 125

```
Tyr Asp His Met Ser Ala Glu Glu Lys Ser Ala Met Arg Arg Ala Glu
    130                 135                 140

Glu Ala Ile Leu Ala Asp Val Asp Val Pro Arg Asp Thr Ala Pro Ser
145                 150                 155                 160

Lys Thr Ala Ala Val Leu Asp Glu Ala Asp Val Gln Ala Leu Ile Asp
                165                 170                 175

Ala Ala Arg Ala Ala Arg Leu Gln Ala Tyr Ala Pro Tyr Ser Gly Tyr
            180                 185                 190

Ala Val Gly Ala Ala Leu Leu Ala Ala Asp Gly Arg Arg Phe Cys Gly
        195                 200                 205

Val Asn Val Glu Asn Ala Ser Tyr Gly Ala Thr Cys Cys Ala Glu Arg
    210                 215                 220

Thr Ala Leu Phe Ala Ala Val Thr Ala Gly Ala Arg Asp Phe Ile Ala
225                 230                 235                 240

Leu Ala Leu Val Thr Glu Gly Asp Glu Pro Ala Pro Pro Cys Gly Leu
                245                 250                 255

Cys Arg Gln Ala Leu Ala Glu Phe Ser Pro Asp Leu Ala Ile Tyr Leu
            260                 265                 270

Ala Gly Pro Thr Gly Glu Thr Tyr Arg Arg Thr Ser Leu Ala Ala Leu
        275                 280                 285

Phe Pro Glu Ala Phe Ser Leu Ser Thr Lys Glu Ser Val
    290                 295                 300
```

<210> 122
<211> 528
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.13;
tr|F2NP91|F2NP91_MARHT polypeptide sequence

<400> 122

```
Met Pro Val Met Glu Thr His Ala Leu Glu Ala Arg Phe Lys Glu Ala
1               5                  10                  15

Leu Ala Arg Leu Cys Pro Glu Gly Arg Leu Leu Ala Ala Val Ser Gly
            20                  25                  30

Gly Gly Asp Ser Val Ala Leu Leu Tyr Leu Leu Lys Ala Ala Gly Arg
        35                  40                  45

Asp Thr Ile Val Ala His Leu Asp His Ala Leu Arg Pro Asp Ser Ala
    50                  55                  60

Ala Asp Ala Ala Phe Val Glu Lys Leu Ala Gln Arg Leu Gly Phe Pro
65                  70                  75                  80

Leu Glu Thr Glu His Val Asp Val Arg Ala Leu Ala His Arg Lys Arg
                85                  90                  95

Ile Asn Leu Glu Ala Ala Ala Arg Glu Val Arg Tyr Ala Phe Leu Ala
            100                 105                 110

Arg Val Ala Arg Arg Trp Lys Ala Arg Cys Ile Leu Thr Ala His Thr
        115                 120                 125

Leu Asp Asp Asn Ala Glu Thr Val Leu Leu Gln Ile Leu Arg Gly Ala
    130                 135                 140

Gly Arg Gly Leu Gly Ile Arg Pro Leu Gln Arg Arg Val Ala Arg Pro
145                 150                 155                 160

Leu Leu Glu Phe Ser Arg Ala Glu Leu Arg Ala Tyr Leu Glu Ala Arg
                165                 170                 175

Gly Ala Arg Trp Leu Glu Asp Pro Thr Asn Arg Ser Leu Glu Leu Asp
            180                 185                 190

Arg Asn Tyr Leu Arg His Ala Val Leu Pro Arg Ile Thr Ala Arg Phe
        195                 200                 205
```

```
Pro His Ala Leu Glu Ala Leu Ala Arg Phe Ser Gln Ala Gln Gln Ala
    210                 215                 220

Asp Asp Trp Ala Leu Glu Ala Leu Ser Ala Arg His Leu Ile Pro Asp
225                 230                 235                 240

Arg Arg Trp Pro Val Pro Ala Tyr Arg Ala Leu Pro Leu Glu Arg Ala
                245                 250                 255

Pro Glu Ala Leu Arg Arg Arg Ala Ile Arg Gly Val Leu Glu Ala Leu
            260                 265                 270

Gly Val Arg Pro Glu Ala Arg Leu Val Ala Asp Val Glu Ala Ala Leu
        275                 280                 285

Gly Gly Arg Ala Gln Thr Leu Pro Gly Gly Val Val Val Arg Arg Gln
    290                 295                 300

Arg Gly Thr Leu Phe Phe Ile Pro Pro Thr Val Arg Phe Pro Lys Val
305                 310                 315                 320

Gln Pro Pro Ala Gly Leu Glu Ala Arg Pro Pro Arg Pro Gly Asp Tyr
                325                 330                 335

Leu Val Phe Pro Tyr Gly Arg Lys Arg Leu Val Asp Phe Leu Asn Glu
            340                 345                 350

Arg Gly Val Pro Arg Glu Leu Lys Arg Arg Trp Pro Val Gly Ala Val
        355                 360                 365

Gly Ala Glu Val Arg Trp Val Tyr Gly Leu Trp Pro Glu Pro Asp Glu
    370                 375                 380

Asp Arg Tyr Met Arg Arg Ala Leu Val Leu Ala Arg Ala Ala Ala Arg
385                 390                 395                 400

Gln Gly Glu Val Pro Ile Gly Ala Val Leu Val Arg Asp Gly Ala Val
                405                 410                 415

Leu Ala Glu Ala Ala Asn Ala Val Glu Ala Ser Arg Asp Ala Thr Ala
```

```
            420                 425                 430

His Ala Glu Leu Leu Ala Leu Arg Thr Ala Leu Arg Arg Val Gly Glu
        435                 440                 445

Lys Val Leu Pro Gly Ala Thr Leu Tyr Val Thr Leu Glu Pro Cys Pro
    450                 455                 460

Met Cys Tyr Gly Ala Ile Leu Glu Ala Arg Val Ala Arg Val Val Tyr
465                 470                 475                 480

Gly Val Glu Asn Leu Lys Ala Gly Ala Phe Thr Val His Gly Leu Glu
                485                 490                 495

Pro Arg Val Ala Leu Glu Ala Gly Arg Val Glu Gly Glu Cys Ala Lys
            500                 505                 510

Val Leu Lys Asp Phe Phe Ala Arg Leu Arg Pro Gly Arg Asp Gly Ala
        515                 520                 525


<210>  123
<211>  226
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.14;
       tr|A0A316TX77|A0A316TX77_9BACT polypeptide sequence

<400>  123

Met Ile Asn Gly Tyr Thr Pro Tyr Ser Gly Asn Gln Asn Thr Cys Tyr
1               5                   10                  15

Val Lys Gly Glu Ser Gly Thr Phe Tyr Pro Gly Val Arg Ile Glu Asn
            20                  25                  30

Val Ser Tyr Pro Leu Thr Ile Ser Ser Val Gln Ala Ala Val Cys Ser
        35                  40                  45

Cys Leu Ala Asn Ser Asp Asn Pro Val Glu Tyr Tyr Thr Gly Asp His
    50                  55                  60
```

```
Gln Pro Glu Leu Leu Gln Val Trp Ala Asp Glu Tyr Asp Met Lys Pro
65                  70                  75                  80

Gly Gly Lys Leu Pro Asp Ser Pro Leu Lys Leu Phe Asp Pro Leu Val
                85                  90                  95

Pro Ser Ile Pro Asp Ile Lys Lys Glu Leu Asp Val Leu Thr Glu Lys
            100                 105                 110

Ser Val Thr Pro Asn Ser Gly Phe Pro Val Ser Ala Leu Leu Gln Thr
        115                 120                 125

Glu Lys Gly Tyr Ile Arg Gly Val Asn Ile Glu Leu Ser Ser Trp Ala
    130                 135                 140

Leu Gly Leu Cys Ala Glu Arg Val Ala Ile Ser Arg Ala Leu Thr Ala
145                 150                 155                 160

Gly Tyr Thr Gln Phe Lys Ser Ile His Ile Tyr Ala Pro Glu Ala Asp
                165                 170                 175

Phe Val Ser Pro Cys Gly Ala Cys Arg Gln Val Leu Leu Glu Val Met
            180                 185                 190

Pro Asp Ala Asp Thr Glu Leu Tyr His Gly Asp Gly Thr Leu Ser Lys
        195                 200                 205

His Ile Val Ser Asp Leu Leu Pro Phe Gly Phe Thr Ser His Lys Leu
    210                 215                 220

Lys Lys
225
```

<210> 124
<211> 321
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.15; tr|R6VYG3|R6VYG3_9FIRM polypeptide sequence

<400> 124

```
Met Ile His Lys Gly Thr Gln Thr Ile Glu Thr Lys Arg Leu Ile Leu
1               5                  10                  15

Arg Ala Phe Thr Pro Asp Asp Ala Glu Ala Ala Phe Glu Asn Trp Met
            20                  25                  30

Ser Asp Pro Lys Val Thr Glu Phe Leu Arg Trp Lys Thr His Ala Asp
        35                  40                  45

Ile Ser Asp Ser Arg Lys Ile Val Asn Glu Trp Ala Asn Gly Ser Ala
    50                  55                  60

Asp Pro Glu Phe Tyr Gln Trp Ala Ile Val Pro Lys Asp Val Asn Glu
65                  70                  75                  80

Pro Ile Gly Thr Ile Ser Val Val Asp Arg Asn Asp Ala Leu Gly Ile
                85                  90                  95

Phe His Ile Gly Tyr Cys Ile Gly Ser Lys Trp Trp His Lys Gly Ile
            100                 105                 110

Thr Ser Glu Ala Phe Ser Ala Val Ile His Phe Leu Phe Glu Glu Val
        115                 120                 125

Gly Ala Asn Arg Ile Glu Ser Gln His Asp Pro Glu Asn Ile His Ser
    130                 135                 140

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

Gln Ala Asp Phe Asn Asn Arg Gly Ile Val Asp Ala Cys Val Tyr Ser
                165                 170                 175

Ile Leu Gln Ser Glu Trp Gln Asn Asn Thr Ser Val Trp Gln Arg Leu
            180                 185                 190

Tyr Asn Ala Ala Leu Thr Val Gln Asn Asp Arg Val Val Ser Pro Phe
        195                 200                 205

Ile Asp Ala Gly Gly Val Ala Ala Ala Leu Met Thr Lys Lys Gly Asn
```

```
    210                 215                 220

Ile Tyr Thr Gly Ile Cys Ile Asp Thr Ala Ser Thr Leu Gly Met Cys
225                 230                 235                 240

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

Ile Asp Lys Ile Val Ala Val Met Pro Asp Gly Lys Val Gly Ala Pro
            260                 265                 270

Cys Gly Ala Cys Arg Glu Tyr Met Met Gln Leu Asp Arg Asp Ser Gly
        275                 280                 285

Asp Ile Glu Ile Leu Leu Asp Leu Glu Thr Glu Lys Thr Val Arg Leu
    290                 295                 300

Lys Asp Leu Ile Pro Asp Trp Trp Gly Ala Glu Arg Phe Gly Asp Thr
305                 310                 315                 320

Glu
```

<210> 125
<211> 229
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.16;
       tr|A0A3C1HZ18|A0A3C1HZ18_9BACI polypeptide sequence

<400> 125

```
Met Gly Asp Ile Met Glu Asn Trp Asn Glu Leu Ser Glu Pro Trp Lys
1               5                   10                  15

Arg Cys Phe Leu Gln Ala Trp Lys Ala Tyr Cys His Gly Ser Ile Pro
            20                  25                  30

Ile Gly Ala Val Leu Val Asp Ser Glu Gly Glu Ile Phe Leu Glu Gly
        35                  40                  45
```

```
Arg Asn Arg Val His Glu Leu Thr Ala Pro Glu Gly Gln Leu Cys Asp
    50                  55                  60

Cys Arg Ile Ala His Ala Glu Met Asn Val Leu Val Gln Val Lys Thr
65                  70                  75                  80

Ser Asp Tyr Glu Lys Leu Ser Gly Ala Thr Ile Tyr Ser Thr Met Glu
                85                  90                  95

Pro Cys Ile Gln Cys Phe Gly Ala Ile Ile Leu Ser Arg Ile Lys Asn
            100                 105                 110

Ile Ser Phe Ala Ala Ile Asp Asp Lys Leu Ala Gly Ala Thr Thr Leu
        115                 120                 125

Glu Asp Arg His Gly Phe Ile Lys Ser Arg Asn Leu Asn Ile Ala Gly
    130                 135                 140

Pro Phe Ser His Leu Gly Glu Ile Gln Ile Ile Leu Arg Thr Asp Phe
145                 150                 155                 160

Leu Leu Arg Ile Phe Asp Ser Glu Tyr Ala Asp Pro Leu Ile Ala Ala
                165                 170                 175

His Glu Lys Asp Tyr Pro Ile Gly Val Ala Leu Gly Arg His Tyr His
            180                 185                 190

Arg Asn Asn Arg Leu Gln Val Ala Lys Lys Glu Thr Ile Pro Phe Gly
        195                 200                 205

Glu Leu Phe Asn Glu Phe Ser Phe Asp Ile Lys Arg Ala Arg Glu Gly
    210                 215                 220

Tyr Thr Leu Gly Lys
225
```

<210> 126
<211> 205
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of artificial sequence: pYY-BEM4.17;
tr|A0A1M6KV24|A0A1M6KV24_9BACT polypeptide sequence

<400> 126

```
Met Glu Ala Ser Gln Gln Asn Ile Leu Leu Lys Ile Glu Gly Lys Gly
1               5                  10                  15


Pro Val Ala Glu Ile Asn Phe Thr Val Thr Leu Pro Glu Trp Leu Val
            20                  25                  30


Glu Gln Val Gln Ser Gly Ser Thr Val Phe Leu Thr Gln Lys Glu Lys
        35                  40                  45


Met Arg Phe Val Leu Glu Leu Ala Arg Lys Asn Val Ala Gln Glu Thr
    50                  55                  60


Gly Gly Pro Phe Ala Ala Ala Val Phe Ser Leu Glu Ser Gly Glu Leu
65                  70                  75                  80


Val Ser Ala Gly Val Asn Val Val Val Glu Ser Arg Cys Ser Ser Ala
                85                  90                  95


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


His Asp Leu Gly Ala Ala Gly Leu Pro Arg Met Val Leu Val Ser Ser
        115                 120                 125


Ala Glu Pro Cys Ala Met Cys Met Gly Ala Ile Pro Trp Ser Gly Val
    130                 135                 140


Lys Gln Val Ile Cys Gly Ala Arg Asp Glu Asp Val Arg Ser Val Gly
145                 150                 155                 160


Phe Asp Glu Gly Ala Lys Pro Leu Glu Trp Val Glu Asp Phe Ala Glu
                165                 170                 175


Arg Gly Ile Glu Val Ile Arg Asp Val Leu Arg Glu Glu Ala Thr Glu
            180                 185                 190


Val Leu Trp Asp Tyr Arg Glu Arg Gly Gly Glu Ile Tyr
```

195 200 205

<210> 127
<211> 207
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.18;
tr|A0A2U0T9B4|A0A2U0T9B4_9RHIZ polypeptide sequence

<400> 127

Met Glu Thr Ala Glu Leu Ile Ser Arg Leu Leu Asp Val Ile Glu Lys
1 5 10 15

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

Phe Gly Ala Ala Ile Leu Lys Lys Ser Asp Leu Ala Val Ile Val Ala
35 40 45

Glu Thr Asn Asn Glu Ile Glu Asn Pro Leu Trp His Gly Glu Met Gln
50 55 60

Ala Ile Lys Arg Phe Phe Glu Leu Pro Ala Asp Gln Arg Pro Ala Thr
65 70 75 80

Arg Asp Cys Leu Phe Leu Ala Thr His Glu Pro Cys Ser Leu Cys Leu
85 90 95

Ser Gly Ile Thr Trp Ser Gly Phe Asp Asn Phe Tyr Tyr Leu Phe Ser
100 105 110

His Gln Asp Ser Arg Asp Gly Phe Ala Ile Pro Tyr Asp Ile Gln Ile
115 120 125

Leu Lys Ser Val Tyr Ala Val Pro Glu Pro Glu Thr Gly Thr Val Ser
130 135 140

Pro Ala Arg Asp Leu Tyr Asn Arg Ser Asn Asp Phe Trp Thr Ser His
145 150 155 160

```
Gly Leu Gln Asp Met Ile Ala Gly Leu Ala Arg Ser Asn Arg Glu Ala
                165                 170                 175


Leu Leu Ala Arg Ile Asp Asp Leu Asn Ala Leu Tyr Ala Glu Leu Ser
            180                 185                 190


Glu Arg Tyr Gln Arg Asp Lys Gly Gly Lys Gly Ile Pro Leu Pro
        195                 200                 205


<210>  128
<211>  231
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.19;
       tr|A0A2K9PN08|A0A2K9PN08_9FLAO polypeptide sequence

<400>  128

Met Ser Asp Lys Lys Glu Ser Lys Ile Lys Ile Ser Lys Thr Ser Glu
1               5                   10                  15


Ser Ile Glu Leu Asp Glu Ile His Ser Leu Leu Ser Tyr Ser Ile Val
            20                  25                  30


Gln Lys Phe Trp Glu Asn Asp Asp Arg Asn Gly Arg Gly Tyr Asn Val
        35                  40                  45


Gly Val Ile Leu Val Asp Glu Asn Lys Asn Ile Val Asp Trp Asp Ile
    50                  55                  60


Asn Ser Val Asn Lys Thr Glu Asn Ser Thr Gln His Gly Glu Met Arg
65                  70                  75                  80


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


Tyr Thr Met Tyr Pro Thr Leu Glu Pro Cys Ala Met Cys Ala Gly Met
            100                 105                 110


Met Thr Met Thr Asn Val Tyr Arg Thr Val Asn Gly Gln Met Asp Tyr
        115                 120                 125
```

```
Phe Tyr Ser Lys Ala Leu Glu Arg Leu Ser Ile Asp Thr Arg Glu Cys
    130                 135                 140

Gly Gly Tyr Pro Pro Tyr Pro Arg Thr Val Ile Ser Glu Ile Ser Pro
145                 150                 155                 160

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

Ala Gly Asn Lys Pro Ile Ile Thr Lys Phe Leu Ser Thr Tyr Lys Ala
            180                 185                 190

Lys Thr Ile Tyr Asp Asp Ala Phe Asn Gln Phe Ile Asn Phe Lys Cys
        195                 200                 205

Lys Phe Pro Glu Asn Lys Thr Lys Tyr Glu Asn Ala Ile Lys Phe Tyr
    210                 215                 220

Asn Ser Leu Pro Glu Ser Ile
225                 230
```

<210> 129
<211> 330
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.20;
tr|F4PWM7|F4PWM7_CAVFA polypeptide sequence

<400> 129

```
Met Arg Phe Ser Leu Ser Leu Leu Phe Val Ile Leu Ser Val Leu Leu
1               5                   10                  15

Ala Gly Val Leu Ala Cys Lys Asp Pro Tyr Asn Pro Glu Thr Val Asp
            20                  25                  30

Tyr Gly Gln Cys Ala Ser Ala Thr Lys Ala Asn Tyr Glu Val Arg Ser
        35                  40                  45

Asp Ser Lys Val Leu Thr Pro Ala Asp Leu Pro Ala Asp Glu Leu Ala
    50                  55                  60
```

```
Val His Glu Ser Arg Met Arg His Ile Ile Asp Ile Ala Arg Val Asn
65                  70                  75                  80

Asn Lys Lys Phe Val Ser Ser Ile Tyr Phe Pro Asn Gly Thr Leu Ala
                85                  90                  95

Cys Ile Gly Ile Asn Thr Gly Lys Pro Asn Met Ile Ala His Gly Glu
            100                 105                 110

Ile Val Ala Ile Gln Asn Cys Thr Glu Ile His Gly Ile Ser Met Tyr
        115                 120                 125

Thr Asn Tyr Ser Ile Tyr Thr Thr Gly Glu Pro Cys Ser Met Cys Ala
    130                 135                 140

Ser Ala Ile Leu Trp Ser Arg Phe Lys Thr Val Val Trp Ser Thr Tyr
145                 150                 155                 160

Asn Ser Asp Leu Tyr Cys Lys Ile Cys Met Ser Asn Ile Pro Ile Asp
                165                 170                 175

Ser Ser Tyr Ile Phe Ser Arg Ala Tyr Gly Leu Gly Ile Glu Ala Pro
            180                 185                 190

Val Ala Ile Gly Gly Val Val Lys Ala Glu Gly Asp Ala Trp Phe Gly
        195                 200                 205

Thr Tyr Cys Asn Arg Pro Thr Ser Ile Tyr Tyr Ile Ala Pro Lys Cys
    210                 215                 220

Ala Cys Gln Asp Pro Ala Lys Val Ser Pro Leu Lys Phe Thr Gln Thr
225                 230                 235                 240

Arg Thr Thr Val Trp Val Glu Gly Gly Asp Lys Val Val Thr Gln Trp
                245                 250                 255

Asn Ala Ile Ile Ser Asn Pro Ser Asn Ser Thr Ile Val Asp Pro Pro
            260                 265                 270
```

```
Ile Val Ile Ser Pro Ser Val Val Phe Lys Gly Ala Pro Trp Gly Ile
        275                 280                 285


Ser Ala Ala Ser Glu Pro Asn Thr Tyr Lys Leu Ser Tyr Asn Lys Val
    290                 295                 300


Leu Phe Pro Gly Gln Thr Phe Ser Phe Gly Tyr Ser Val Tyr Gly Leu
305                 310                 315                 320


Glu Glu Val Ala Phe Thr Ala Leu Glu Ala
                325                 330


<210>  130
<211>  193
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.21;
       tr|U7QZM1|U7QZM1_PHOTE polypeptide sequence

<400>  130

Met Asn Lys Thr Arg Arg Lys Leu Leu Ala Thr Leu Gly Ile Met Ser
1               5                   10                  15


Ile Ser Met Ser Phe Ile Ala Gln Ala Gly Glu Lys Lys Thr Gln Val
            20                  25                  30


Ile Asn Asn Ile Leu Ser Lys Gln Glu Ile Thr Glu His Glu Lys Tyr
        35                  40                  45


Met Arg Glu Ala Ile Lys Glu Ala Ile Lys Asn Pro Lys His Pro Phe
    50                  55                  60


Gly Ala Val Ile Val Asn Arg Asn Asn Gly Glu Ile Leu Ser Arg Gly
65                  70                  75                  80


Val Asn Thr Gly Arg Asn Asn Pro Ile Leu His Gly Glu Ile Gln Ala
                85                  90                  95


Ile Asn His Tyr Ile Thr Gln Tyr Gly Asn Gln Gly Trp Glu Asn Val
            100                 105                 110
```

```
Ala Leu Tyr Thr Thr Gly Glu Pro Cys Ser Met Cys Met Ser Ala Leu
        115                 120                 125


Val Trp Ile Gly Ile Arg Glu Val Ile Trp Ala Thr Ser Ile Ser Val
    130                 135                 140


Ile Arg Asn Ser Gly Ile Arg Gln Ile Asp Ile Ser Ala His Glu Ile
145                 150                 155                 160


Ala Glu Arg Ala Ser Ser Phe Tyr Asn Pro Ile Thr Leu Val Gly Gly
                165                 170                 175


Ile Leu Ala Asn Glu Thr Asp Lys Leu Phe Leu Glu Arg Lys Arg Gly
            180                 185                 190


Asn
```

<210> 131
<211> 319
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.22;
tr|A0A081CH48|A0A081CH48_PSEA2 polypeptide sequence

<400> 131

```
Met Ala Ser Arg Arg His Leu Leu Ala Thr Gln Val Thr Gly Asn His
1               5                   10                  15


Arg Lys Leu Ser Leu Trp His Leu Arg Gly Trp Leu Ser Pro Tyr Thr
            20                  25                  30


Lys Leu Val Asp Ala Val Tyr Phe Leu Thr Thr Asn Ser Phe Tyr His
        35                  40                  45


Ser Leu Gln Thr Pro Pro Val Gln Ser Ile Thr Met Leu Leu Ser Ser
    50                  55                  60


Ile Ile Thr Ser Leu Ala Leu Ala Ala Gln Ala Ser Ala Tyr Arg Glu
65                  70                  75                  80
```

```
Gly Leu His Pro Glu Phe Gln Ser Gly Leu Ser Ile Asn Ser Val Pro
                85                  90                  95

Ala Thr Asp Arg Asp His Trp Met Arg Leu Ala Asn Ser Ala Ile Tyr
            100                 105                 110

Tyr Pro Pro Val Ser His Pro Cys Pro Gln Ala Pro Phe Gly Thr Ala
        115                 120                 125

Ile Val Asn Thr Thr Ser Asn Glu Leu Ile Cys Ala Ile Ala Asn Arg
    130                 135                 140

Val Gly Ser Thr Gly Asp Pro Thr Gln His Gly Glu Ile Thr Ala Ile
145                 150                 155                 160

Gln His Cys Thr Asn Val Met Arg Lys Lys Gly Leu Ser Pro Gln Glu
                165                 170                 175

Ile Ile Ala Ala Trp Lys Gln Leu Ser Leu Tyr Thr Asn Ala Glu Pro
            180                 185                 190

Cys Thr Met Cys Leu Ser Ala Ile Arg Trp Ala Gly Phe Lys Glu Val
        195                 200                 205

Ile Tyr Gly Thr Ser Val Gly Thr Ile Ser Glu Asn Gly Arg Asn Gln
    210                 215                 220

Ile Tyr Ile Pro Ser Asn Leu Val Leu Glu Lys Ser Tyr Ser Phe Gly
225                 230                 235                 240

His Ala Thr Leu Met Leu Gly Asn Ile Leu Thr His Glu Thr Asp Pro
                245                 250                 255

Phe Phe Gln His Gln Phe Asn Glu Ser Ala Pro Cys Pro Val Gly Cys
            260                 265                 270

Glu Arg Thr Gln Val Gly Glu Ala Arg Val Lys Thr Cys Glu Pro Val
        275                 280                 285
```

```
Pro Asn Trp Gln Lys Leu Val Arg Leu Glu Tyr Ser Glu Asp Ser Arg
    290                 295                 300


Val Gly Ser Glu Pro Val Ala His Thr Pro Leu His Leu Glu Leu
305                 310                 315


<210>  132
<211>  203
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.23;
       tr|A0A3D3HMU1|A0A3D3HMU1_9GAMM polypeptide sequence

<400>  132

Met Asp Tyr Ser Asp Ala Ile Leu Gly Ala Ile Thr Ser Ile Arg Arg
1               5                   10                  15


Asn Ser Lys Gln Pro Gly Val Asn Val Thr Asp Asn Val Thr Asp Ser
            20                  25                  30


Ser Thr Gln Tyr Asn Asn Asp Glu Tyr Trp Met Arg Arg Ala Leu Ala
        35                  40                  45


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


Leu Val Lys Asp Asn Gln Gln Val Ala Gly Gly Phe Asn Gln Pro Ile
65                  70                  75                  80


Arg Ser His Asp Pro Ala Ala His Ala Glu Ile Leu Thr Leu Arg Glu
                85                  90                  95


Ala Gly Ala Val Leu Gly Asn Tyr Arg Leu Ile Asp Thr Thr Leu Tyr
            100                 105                 110


Val Thr Leu Glu Pro Cys Met Met Cys Ala Gly Ala Leu Val His Ser
        115                 120                 125


Arg Ile Lys Arg Leu Val Phe Gly Ala Ala Glu Pro Lys Thr Gly Ala
    130                 135                 140
```

Ala Gly Ser Phe Ile Asp Leu Leu Thr Leu Pro Arg Leu Asn His Tyr
145 150 155 160

Met Glu Val Thr Gly Gly Val Leu Gly Glu Glu Cys Ser Val Leu Leu
165 170 175

Ser Asp Phe Phe Arg Arg Arg Arg Ala Glu Lys Lys Ala Leu Lys Arg
180 185 190

Gln Asn Ser Glu Ser Gly Ser Asp Ser Ala Ser
195 200

<210> 133
<211> 270
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.24;
tr|A0A1N5WT13|A0A1N5WT13_9ACTN polypeptide sequence

<400> 133

Met Leu Glu Lys Ile Glu Arg Arg Leu Val Ala Ala Ala Glu Ala Val
1 5 10 15

Val Arg Ser Pro Ser Thr Gly Asp Ala His Thr Val Ala Ala Ala Ala
20 25 30

Met Asp Ala Asn Gly Asp Ile Tyr Ser Gly Val Asn Val Phe His Phe
35 40 45

Thr Gly Gly Pro Cys Ala Glu Leu Val Val Ile Gly Ser Ala Ala Ala
50 55 60

Ala Asn Ala Pro Pro Leu Ile Thr Ile Val Ala Val Gly Asp Gly Asp
65 70 75 80

Arg Gly Val Ile Ala Pro Cys Gly Arg Cys Arg Gln Val Met Leu Asp
85 90 95

Leu His Pro Asp Val Phe Val Ile Val Pro Thr Gly Asp Gly Gln Leu
100 105 110

```
Ala Ala Lys Pro Val Arg Glu Leu Leu Pro Phe Gly Tyr Val Ala Arg
        115                 120                 125

Thr Gly Ser Thr Ala Pro Arg Val Val Tyr Phe His Pro Arg His Tyr
    130                 135                 140

Asp Thr Ile Ser Ser Gly Leu Lys Thr Ala Thr Val Arg Phe Gln Asp
145                 150                 155                 160

Ser Val Gln Thr Gly Pro Ala Val Phe Val Phe Asp Asp Gly Glu Ser
                165                 170                 175

Ile Arg Arg Leu Asp Ala Val Val Glu Lys Val Glu Ser Arg Arg Leu
            180                 185                 190

Asp His Leu Thr Glu Glu Asp Ala His His Glu Ala Leu Pro Asp Ser
        195                 200                 205

Asp Ala Leu Arg Asp Ala Ile Lys Thr Gln Tyr Pro Met Leu Gly Asp
    210                 215                 220

Gly Asp Val Val Asp Val Ala Thr Phe Arg Leu Thr Ala Ile Ser Ala
225                 230                 235                 240

Pro Asp Pro Asp Pro Arg Ser Ser Tyr Pro Pro Ala Val Ser Arg Cys
                245                 250                 255

Asn Pro Ala Gly Pro Arg Ala Asp Leu Leu Val Gly Gln Ser
            260                 265                 270
```

<210> 134
<211> 257
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.25;
tr|X0SAC5|X0SAC5_9ZZZZ polypeptide sequence

<400> 134

```
Met Thr Lys Asp Gly Arg Val Ile Ala Ser Ala His Asp Thr Glu Val
```

```
1               5                   10                  15

Thr Asp Gln Asp Ser Thr Ala His Ala Glu Ile Asn Ala Ile Arg Lys
            20                  25                  30

Ala Ser Lys Ile Tyr Arg Lys Asp Leu Thr Gly Cys Leu Ile Ile Ser
        35                  40                  45

Thr His Glu Pro Cys Pro Met Cys Thr Gly Ser Ile Ile Trp Ser Asn
    50                  55                  60

Ile Ser Lys Val Val Tyr Gly Val Ser Ile Arg Asp Ser Ile Lys Ala
65                  70                  75                  80

Gly Arg Asp Met Ile Asn Leu Ser Cys Lys Glu Ile Ile Lys Lys Pro
                85                  90                  95

Asn Ala Glu Ile Asn Ile Tyr Asp Gly Ile Leu Lys Lys Glu Cys Leu
            100                 105                 110

Lys Leu Tyr Asn Asn Asp Thr Arg Lys Leu Val Lys Lys Phe Arg Lys
        115                 120                 125

Tyr Glu Trp Ile Asn Ile Glu Glu Asn Leu Leu Asn Lys Arg Met Gln
    130                 135                 140

Trp Phe Glu Asn Asn Lys Thr Met Ile Arg Lys Leu Lys Gly Asn Asp
145                 150                 155                 160

Leu Glu Lys Ala Tyr His Leu Ile Leu Met Lys Ile Gly Ile Lys Arg
                165                 170                 175

Ser Glu Ala Pro Ile Val Lys Lys Ser Glu Ser Lys Ile Ile Phe His
            180                 185                 190

Ser Lys Asn Tyr Cys Pro Ser Leu Glu Ala Cys Ile Ile Leu Asp Leu
        195                 200                 205

Asp Thr Arg Glu Val Cys Lys Glu Ile Tyr Glu Arg Pro Thr Glu Glu
    210                 215                 220
```

Leu Ile Arg Arg Leu Asn Ser Lys Leu Arg Phe Thr Arg Asn Tyr Asp
225 230 235 240

Cys Ile Arg Pro Tyr Ser Asp Tyr Cys Glu Glu Ile Ile Ile Leu Glu
245 250 255

Lys

<210> 135
<211> 154
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.26;
tr|A0A3B8IC10|A0A3B8IC10_9BACT polypeptide sequence

<400> 135

Met Pro Ser His Glu Asp Phe Ile His Gln Cys Leu Glu Leu Gly Lys
1 5 10 15

Glu Ala Leu Leu Gln Gly Asn Pro Pro Val Gly Ser Val Ile Val Trp
20 25 30

Gln Asp Gln Val Ile Gly Arg Gly Ile Glu Asn Gly Arg Ser Ser Gly
35 40 45

Asp Ile Thr Gln His Ala Glu Leu Leu Ala Leu Gln Glu Ala Val Ala
50 55 60

Thr Gly Gln Arg Asp Lys Leu Lys Glu Ala Ile Ile Tyr Ser Thr His
65 70 75 80

Glu Pro Cys Val Met Cys Ala Tyr Pro Ile Arg Gln Tyr Lys Ile Pro
85 90 95

Thr Val Val Tyr Ser Val Ala Val Pro Glu Leu Gly Gly His Thr Ser
100 105 110

Ser Trp His Leu Leu Thr Thr Glu Asp Val Pro Lys Trp Gly Lys Ala
115 120 125

Pro Lys Ile Ile Thr Gly Ile Ser Ala Glu Glu Val Glu Ala Leu Asn
130 135 140

Ala Ala Phe Gln Asp Ser Leu Lys Lys Gly
145 150

<210> 136
<211> 180
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.27;
tr|A0A2N9P8B9|A0A2N9P8B9_9FLAO polypeptide sequence

<400> 136

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

Gln Asp Lys Ile Ile Gln Lys Leu Tyr Ile Cys Phe Met Glu Asn Ile
20 25 30

Phe Thr Asp Glu Tyr Phe Met Lys Lys Ala Leu Gln Glu Ala Glu Thr
35 40 45

Ala Phe Gln Gln Gly Glu Ile Pro Val Gly Ala Val Ile Val Ile Asp
50 55 60

Asn Arg Ile Ile Ala Arg Ser His Asn Leu Thr Glu Met Leu Asn Asp
65 70 75 80

Val Thr Ala His Ala Glu Met Gln Ala Ile Thr Ala Ser Ala Asn Phe
85 90 95

Leu Gly Gly Lys Tyr Leu Lys Asp Cys Thr Leu Tyr Val Thr Leu Glu
100 105 110

Pro Cys Gln Met Cys Ala Gly Ala Leu Tyr Trp Ser Gln Ile Ser Lys
115 120 125

Ile Val Tyr Gly Ala Thr Asp Glu Gln Arg Gly Tyr Arg Ala Met Gly

130 135 140

Ala Gln Leu His Pro Lys Thr Lys Val Ile Ser Gly Ile Met Gln Asn
145 150 155 160

Glu Cys Thr His Leu Met Lys Asp Phe Phe Lys Gln Arg Arg Ser Lys
165 170 175

Ser Thr Lys Asp
180

<210> 137
<211> 178
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.28;
tr|K1KX30|K1KX30_9BACT polypeptide sequence

<400> 137

Met Val Lys Asn Pro Val Asn Asn Asn Glu Leu Tyr Phe Gly Lys His
1 5 10 15

Ser Glu Ile Pro Met Asn Glu Glu Gln Lys Ala Tyr Met Lys Met Ala
20 25 30

Val Asp Leu Ser Arg Ser Gly Met Glu Ser Gly Lys Gly Gly Pro Phe
35 40 45

Gly Cys Val Ile Val Lys Asp Gly Lys Val Ile Gly Ile Gly Ser Asn
50 55 60

Ser Val Leu Glu Thr Asn Asp Pro Thr Ala His Ala Glu Ile Val Ala
65 70 75 80

Ile Arg Asp Ala Cys Arg Asn Leu Gly His Phe Gln Leu Asp Gly Cys
85 90 95

Glu Val Tyr Thr Ser Cys Glu Pro Cys Pro Met Cys Leu Gly Ala Ile
100 105 110

Tyr Trp Ala Arg Pro Ser Lys Val Phe Phe Ala Asn Asp Lys Arg Asp
        115   120   125

Ala Ala Glu Ala Gly Phe Asp Asp Asp Phe Ile Tyr Gln Glu Leu Glu
    130   135   140

Leu Pro Tyr Glu Lys Arg Lys Ile Pro Phe Glu Gln Gly Met Gln Asp
145   150   155   160

Thr Ala Lys Glu Val Phe Gln Glu Trp Ile Leu Lys Glu Asp Lys Thr
                165   170   175

Leu Tyr

<210> 138
<211> 196
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.29;
       tr|R4XI84|R4XI84_TAPDE polypeptide sequence

<400> 138

Met Ser Ser Glu Ile Glu Pro Pro Ser Thr Asp Val His Lys His Ala
1   5   10   15

Val Ala Glu Ala Ala Asp Glu Ser Gly Ala Ala Asp Ala Phe Met Gln
            20   25   30

Ile Ala Leu Gln Gln Ala Glu Thr Ala Leu Leu Asn Lys Glu Val Pro
        35   40   45

Val Gly Cys Val Phe Val His Gln Pro Thr Gly Thr Val Leu Ala Thr
    50   55   60

Gly Ala Asn Gln Thr Asn Ala Ser Leu Asn Gly Thr Leu His Ala Glu
65   70   75   80

Phe Val Ala Ile Glu Ser Ile Leu Arg Asp His Pro Pro Ser Ile Phe
                85   90   95

```
Arg Glu Ser Asp Leu Tyr Val Thr Val Glu Pro Cys Val Met Cys Ala
            100                 105                 110

Ser Ala Leu Arg Gln Leu Gln Val Arg Lys Val Tyr Phe Gly Cys Gly
        115                 120                 125

Asn Asp Arg Phe Gly Gly Cys Gly Ser Val Phe Ser Ile His Ser Asp
    130                 135                 140

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

Arg Lys Glu Ala Ile Met Leu Leu Arg Arg Phe Tyr Leu Leu Gln Asn
                165                 170                 175

Glu Ser Ala Pro Lys Pro Ala Leu Lys Ser Thr Arg Val Leu Lys Glu
            180                 185                 190

His Phe Asp Glu
        195
```

<210> 139
<211> 212
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.30;
       tr|A0A239CVF7|A0A239CVF7_9DELT polypeptide sequence

<400> 139

```
Met Ser Pro Ala Ser Lys Lys His Phe Pro Ser Leu Phe Ser Phe Leu
1               5                   10                  15

Leu Leu Thr Ile Gly Leu Ile Cys Gly Thr Ala His Ala Gln Pro Gln
            20                  25                  30

Gly His Thr Ala Asp Asp Thr Ala Ala Thr Leu Ala Asn Ala Ser Leu
        35                  40                  45

Lys Glu His Glu Pro Phe Ile Arg Arg Cys Tyr Gln Leu Ala Ile Asp
    50                  55                  60
```

```
Ala Gly Lys Lys Gly Asn His Pro Phe Gly Ala Leu Leu Val His Lys
65                  70                  75                  80

Gly Lys Ile Val Leu Glu Ala Glu Asn Thr Val Leu Thr Asp Asn Asp
                85                  90                  95

Phe Thr Asn His Ala Glu Met Asn Leu Ile Ala Glu Ala Ala Arg Thr
            100                 105                 110

Leu Ser Arg Gln Ile Ile Pro Glu Ala Thr Val Tyr Thr Ser Cys Ala
        115                 120                 125

Pro Cys Ala Met Cys Thr Ala Thr Leu Ala Met Ala Gly Phe Thr Arg
    130                 135                 140

Ile Val Tyr Gly Val Ser His Asp Ala Leu Asn Lys Arg Phe Gly Leu
145                 150                 155                 160

Lys Gly Lys Ser Val Ser Cys Pro Ala Leu Phe Lys Thr Met Gly Met
                165                 170                 175

Glu Leu Glu Phe Val Gly Pro Val Leu Glu Lys Glu Gly Leu Arg Val
            180                 185                 190

Phe Asp Phe Trp Pro Glu Lys Asp Pro His Ala Gln Met Leu Lys Lys
        195                 200                 205

Gln Ala Arg Lys
    210
```

<210> 140
<211> 320
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.31;
       tr|A0A1Q3NME1|A0A1Q3NME1_9BACT polypeptide sequence

<400> 140

```
Met Thr Glu Phe Asn Tyr Asp Trp Ala Lys Leu Ala Phe Ser Ser Lys
```

```
1               5                   10                  15

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

Ile Ser Glu Lys Arg Phe Thr Gln Leu Leu Lys Glu Tyr Leu Pro Lys
        35                  40                  45

Gly Asp Ile Leu Leu Gly Ile Ser Lys Glu Asp Tyr Val Glu Gly Leu
    50                  55                  60

Glu Gly Gln Pro Gln Phe Ala Met Leu Gln Gln Lys Thr Leu Gln Lys
65                  70                  75                  80

Leu Ile Asp Lys Val Asn Asp Ala Ser Ala His Lys Val Tyr Thr Leu
                85                  90                  95

Arg Tyr Phe Gln Arg Glu Leu Pro Ala Ile Ile Glu Lys Leu Thr Pro
            100                 105                 110

Pro Arg Val Val Gly Ile His Gly Ser Trp His His Ser Phe His Thr
        115                 120                 125

Leu Pro Ile Tyr Tyr Leu Leu Ser Glu Lys Arg Ile Pro Tyr Gln Leu
    130                 135                 140

Val Ala Ala Phe Ser Asp Glu Asp Glu Ala Arg Ala Tyr Glu Val Ala
145                 150                 155                 160

Thr Asp Lys Lys Ile Val Arg Pro Thr Leu Glu Gly Ser Phe Asp Asp
                165                 170                 175

Thr Thr Val Leu Gln Leu Thr Asp Glu Val Ala Lys Ser Ser Tyr Asp
            180                 185                 190

Tyr Gly Phe Gln Thr Gly Ala Ile Leu Ala Glu Lys Val Asn Gly Val
        195                 200                 205

Tyr Gln Pro Val Ala Ala Gly Phe Asn Lys Val Val Pro Tyr Gln Thr
    210                 215                 220
```

```
Tyr Ala Leu Leu Asn Gly Ala Ser Arg Glu Thr Asn Phe Ser Pro Ala
225                 230                 235                 240


Asn Asp Met Asn His Tyr Asp Thr Ile His Ala Glu Met Gln Ile Leu
                245                 250                 255


Val Glu Ala Ala Lys Gln Gly Ile Ser Leu Lys Asp Lys Thr Leu Phe
            260                 265                 270


Val Asn Leu Met Pro Cys Pro Ser Cys Ala Arg Thr Leu Ser Gln Thr
        275                 280                 285


Glu Leu Ser Glu Ile Val Tyr Arg Ile Asp His Ser Gly Gly Tyr Ala
    290                 295                 300


Val Asp Leu Leu Thr Lys Val Gly Lys Asp Ile Arg Arg Ile Val Tyr
305                 310                 315                 320


<210>  141
<211>  180
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.32;
       tr|A0A2G6N4N7|A0A2G6N4N7_9DELT polypeptide sequence

<400>  141

Met Lys Glu Arg Thr Val Ser Tyr Ser Asp Arg His Phe Met Ala Glu
1               5                   10                  15


Ala Leu Glu Met Ala Glu Ser Ala Leu Thr Gln Gly Glu Phe Pro Val
            20                  25                  30


Gly Cys Val Ile Ala Asp Gly Thr Ala Val Val Ala Arg Gly His Arg
        35                  40                  45


Thr Gly Thr Thr Ala Gly Ala Val Asn Glu Ile Asp His Ala Glu Ile
    50                  55                  60


Asn Ala Leu Arg His Leu Gly Leu Ala Gly Glu His Leu Asp Arg Thr
65                  70                  75                  80
```

```
Asp Leu Thr Ile Tyr Ser Thr Met Glu Pro Cys Leu Met Cys Phe Ala
                85                  90                  95

Ala Ile Val Leu Ser Gly Ile Asn Arg Ile Val Tyr Ala Tyr Glu Asp
            100                 105                 110

Val Met Gly Gly Gly Thr Gly Cys Asp Leu Thr Gly Leu Pro Pro Leu
        115                 120                 125

Tyr Arg Asp Ala Pro Leu Thr Leu Val Ala Gly Val Arg Arg Arg Ala
    130                 135                 140

Ser Leu Asn Leu Phe Arg Arg Phe Phe Thr Asp Pro Glu Asn Gly Tyr
145                 150                 155                 160

Trp Ala Gly Ser Leu Leu Ser Arg Tyr Thr Leu Asn Gln Thr Lys Asp
                165                 170                 175

Ser His Arg Leu
            180
```

<210> 142
<211> 156
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.33;
tr|A0A0G0RBB8|A0A0G0RBB8_9BACT polypeptide sequence

<400> 142

```
Met Gln Ser Val Gln Tyr Asn Lys Leu Thr His Leu Gln Arg Arg Ala
1               5                   10                  15

Leu Asp Glu Ala Glu Gln Val Leu Glu Asn Ser Tyr Asn Pro Tyr Ser
            20                  25                  30

His Phe Tyr Val Gly Ala Cys Leu Ile Ser Glu Asp Glu Gln Leu Ile
        35                  40                  45

Ala Gly Thr Asn Phe Glu Asn Ala Ala Tyr Gly Ser Ala Ile Cys Ala
```

```
    50                  55                  60

Glu Arg Ala Ala Val Leu Arg Ala Asn Ala Met Ser Ile Arg Arg Phe
65                  70                  75                  80

Arg Gly Ile Ala Ile Ile Ala Arg Gly Glu Asp Phe Asn Thr Thr Glu
                85                  90                  95

Val Thr Gly Pro Cys Gly Ser Cys Arg Gln Val Leu Tyr Glu Ile Ser
            100                 105                 110

Gln Val Ser Gly Cys Asp Leu Gln Val Ile Leu Ala Thr Ser Lys Lys
        115                 120                 125

Asp Lys Ile Val Ile Thr Thr Ile Arg Glu Leu Leu Pro Leu Ala Phe
    130                 135                 140

Gly Pro Leu Asp Leu Gly Val Asp Ile Gly Lys Tyr
145                 150                 155


<210>  143
<211>  217
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.34;
       tr|A0A327L2Q5|A0A327L2Q5_9RHIZ polypeptide sequence

<400>  143

Met Val Thr Ser Arg Asp Gly Glu Asp Glu Ala Met Met Ala Arg Cys
1               5                   10                  15

Val Ala Leu Ser Arg Ile Ala Val Gly Lys Gly Glu Tyr Pro Phe Gly
            20                  25                  30

Ala Val Val Ala Arg Glu Gly Arg Ile Val Ala Glu Ala Ile Asn Arg
        35                  40                  45

Thr Ile Arg Asp Gly Asp Val Ser Arg His Ala Glu Val Ile Ala Leu
    50                  55                  60
```

```
Ala Arg Ala Gln Lys Ala Ile Gly Arg Arg Glu Leu Arg Glu Cys Ser
65                  70                  75                  80


Leu Tyr Ser Asn Val Glu Pro Cys Ala Met Cys Ser Tyr Cys Ile Arg
                85                  90                  95


Glu Ala Trp Val Gly Arg Val Val Tyr Ala Leu Gly Ser Pro Val Met
            100                 105                 110


Gly Gly Val Ser Lys Trp Asn Ile Leu Arg Asp Asp Gly Leu Ser Gly
        115                 120                 125


Arg Met Pro Gln Val Phe Asp Ala Ala Pro Glu Val Val Ser Gly Val
    130                 135                 140


Leu Val Glu Gln Ala Gln Ala Ala Trp Arg Asp Trp Ser Pro Leu Ala
145                 150                 155                 160


Trp Glu Met Ile Thr Leu Arg Gly Leu Met Thr Asp Pro Ser Ala Arg
                165                 170                 175


Pro Glu Cys Arg Thr Arg Ala Ala Arg Pro Arg Ser Leu Trp His His
            180                 185                 190


Leu Val Ala Leu Ile Glu Arg Pro Pro Arg Pro Tyr Val Asp Pro Thr
        195                 200                 205


Ser Ala Ala Glu Gly His Ala Asp Leu
    210                 215
```

<210> 144
<211> 162
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.35;
       tr|S2DR30|S2DR30_9BACT polypeptide sequence

<400> 144

```
Met Lys Met Lys Lys Lys Ile Glu Ile Thr Val Ser Leu Glu Val Ile
1               5                   10                  15
```

```
Gln Lys Ser Glu Trp Ser Lys Glu Asp Arg Ser Leu Ile Glu Arg Ala
            20                  25                  30

Ile His Ala Val Glu His Ala His Ala Pro Tyr Ser Asn Phe Met Val
        35                  40                  45

Gly Thr Ala Leu Leu Leu Asp Asn Gly Gln Ile Phe Ser Ala Asn Asn
    50                  55                  60

Gln Glu Asn Val Ser Phe Pro Val Gly Ile Cys Ala Glu Arg Ala Val
65                  70                  75                  80

Leu Ser Tyr Ala Met Gly Asn Phe Pro Asn Asn Arg Pro Val Lys Leu
                85                  90                  95

Ala Val Val Ala Lys Arg Arg Ser Asp Ser Thr Trp Ala Thr Val Thr
            100                 105                 110

Pro Cys Gly Leu Cys Arg Gln Thr Ile Asn Glu Tyr Glu Val Lys Phe
        115                 120                 125

Gly His Pro Ile Glu Ile Leu Met Leu Asn Pro Gly Glu Glu Ile Leu
    130                 135                 140

Lys Ala Ser Gly Ile Asp Gln Leu Leu Pro Phe Arg Phe Asn Asp Leu
145                 150                 155                 160

Asn Ser
```

<210> 145
<211> 153
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.36; tr|A0A369QGF1|A0A369QGF1_9BACT polypeptide sequence

<400> 145

```
Met Glu Glu His Glu Lys Trp Met His Trp Cys Leu Asn Leu Ala Gln
1               5                   10                  15
```

```
Gln Ala Leu Gln Gln Gly Asp Phe Pro Val Gly Ala Val Val Val Gln
            20                  25                  30

Lys Gly Lys Leu Ile Gly Gln Gly Val Glu Ala Gly Gln Leu Lys Lys
        35                  40                  45

Asp Ile Thr Cys His Ala Glu Met Glu Ala Ile Arg Asp Ala Arg Gln
    50                  55                  60

Thr Ile Asn Thr Ala Asp Leu Gln Asn Cys Ile Leu Tyr Ser Thr His
65                  70                  75                  80

Glu Pro Cys Ile Met Cys Ser Tyr Val Ile Arg His His Lys Ile Ser
                85                  90                  95

Arg Val Val Val Gly Thr Thr Val Pro Glu Val Gly Gly Ser Ser Ser
            100                 105                 110

Ala Tyr Pro Leu Leu Ser Ala Pro Asp Ile Ser Ile Trp Val Ala Pro
        115                 120                 125

Pro His Leu Val Thr Gly Val Leu Ala Glu Ala Cys Gln Ala Leu Ser
    130                 135                 140

Gln Ala Tyr Lys Gln Lys Phe Lys Lys
145                 150
```

<210> 146
<211> 159
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.37;
       tr|A0A1W6X4U4|A0A1W6X4U4_9RHIZ polypeptide sequence

<400> 146

```
Met Thr Asn Pro Ser Arg Gln Glu Arg Trp Asp Arg Arg Phe Leu Glu
1               5                   10                  15

Leu Ala Lys Val Phe Gly Thr Trp Ser Lys Asp Arg Ser Ala Gly Thr
```

```
            20                  25                  30

Gly Cys Val Ile Val Gly Pro Asp Arg Leu Leu Arg Ala Ser Gly Tyr
        35                  40                  45

Asn Gly Phe Ala Arg Gly Ile Asp Asp Glu Val Pro Glu Arg His Glu
    50                  55                  60

Arg Pro Ala Lys Tyr Ser Trp Thr Glu His Ala Glu Arg Asn Ala Ile
65                  70                  75                  80

Tyr Asn Ala Ala Lys Leu Gly Ile Ser Leu Asp Gly Cys Thr Ala Tyr
                85                  90                  95

Val Asn Trp Phe Pro Cys Ile Asp Cys Ala Arg Ala Ile Val Gln Ala
            100                 105                 110

Gly Ile Val Arg Leu Val Gly Leu His Pro Asp His Ala Asp Gln Arg
        115                 120                 125

Trp Gly Ser Glu Phe Lys Phe Ala Thr Glu Met Leu Arg Glu Ser Gly
    130                 135                 140

Ile Glu Ile Ile Leu Tyr Asp Ile Pro Glu Leu Ala Ala Arg Lys
145                 150                 155


<210>  147
<211>  490
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.38;
       tr|A0A238BW09|A0A238BW09_9BILA polypeptide sequence

<400>  147

Met Glu Glu Met Ala Arg Lys Ile Arg Thr Lys Ala Lys Lys Ala Asn
1               5                   10                  15

Ser Tyr Cys Asn Thr Met Thr Phe Leu Ile Ser Lys Ala Ser Ile Val
            20                  25                  30
```

```
Leu Leu Lys Ala Glu Cys Lys Arg Ile Glu Leu Thr Val Val Ile Phe
        35                  40                  45

Arg Phe Leu Ile Lys Met Asn Ala Ser Glu Pro Asn Asn Glu Leu Cys
    50                  55                  60

Asp Met Thr Val Ile Lys Ser Met Leu Lys Ile Thr His Val Ile Phe
65                  70                  75                  80

Asp Leu Asp Gly Leu Leu Ile Asp Thr Glu Val Val Phe Ser Lys Val
                85                  90                  95

Asn Gln Cys Leu Leu Ser Lys Tyr Asn Lys Lys Phe Thr Pro His Leu
            100                 105                 110

Arg Gly Leu Val Thr Gly Met Pro Lys Lys Ala Ala Val Thr Tyr Ile
        115                 120                 125

Leu Glu His Glu Lys Leu Ser Ala Lys Val Asp Val Asp Glu Tyr Cys
    130                 135                 140

Lys Lys Tyr Asp Glu Met Ala Glu Glu Met Leu Pro Lys Cys Ser Leu
145                 150                 155                 160

Met Pro Gly Val Met Lys Leu Val Arg His Leu Lys Thr His Ser Ile
                165                 170                 175

Pro Met Ala Ile Cys Thr Gly Ala Thr Lys Lys Glu Phe Glu Ile Lys
            180                 185                 190

Thr Arg Tyr His Lys Glu Leu Leu Asp Leu Ile Ser Leu Arg Val Leu
        195                 200                 205

Ser Gly Asp Asp Pro Ala Val Lys Arg Gly Lys Pro Ala Pro Asp Pro
    210                 215                 220

Phe Leu Val Thr Met Asp Arg Phe Lys Gln Lys Pro Glu Lys Ala Glu
225                 230                 235                 240

Asn Val Leu Val Phe Glu Asp Ala Ala Asn Gly Val Cys Ala Ala Ile
                245                 250                 255
```

```
Ala Ala Gly Met Asn Val Ile Met Val Pro Asp Leu Thr Tyr Met Lys
            260                 265                 270

Ile Pro Glu Gly Leu Gln Asn Lys Ile Asn Ser Phe Ser Asp Asn Leu
        275                 280                 285

Ile Ile Ser Asn Asp Leu Asn Val Ala Leu Met Ser Leu Lys Lys Glu
    290                 295                 300

Leu Ser Glu Glu Glu Val His Phe Leu Asn Arg Ala Phe Glu Ile Ala
305                 310                 315                 320

Val Asp Ala Val Leu Asn Asn Glu Val Pro Val Gly Cys Val Phe Val
                325                 330                 335

Phe Glu Gly Gln Glu Val Ala Phe Gly Arg Asn Asp Val Asn Arg Thr
            340                 345                 350

Lys Asn Pro Thr Tyr His Ala Glu Met Val Ala Leu Lys Met Met Lys
        355                 360                 365

Gln Trp Cys Met Asp Asn Gly Arg Asp Leu Glu Glu Ile Met Arg Arg
    370                 375                 380

Thr Thr Leu Tyr Val Thr Leu Glu Pro Cys Ile Met Cys Ala Ser Ala
385                 390                 395                 400

Leu Tyr His Leu Arg Leu Lys Lys Ile Leu Tyr Gly Ala Ala Asn Glu
                405                 410                 415

Arg Phe Gly Gly Leu Val Ser Val Gly Thr Arg Glu Lys Tyr Gly Ala
            420                 425                 430

Lys His Phe Ile Glu Ile Met Pro Asn Leu Ser Val Asp Arg Ala Val
        435                 440                 445

Lys Leu Leu Lys Glu Phe Tyr Glu Lys Gln Asn Pro Phe Cys Pro Glu
    450                 455                 460
```

```
Glu Lys Arg Lys Val Lys Lys Pro Lys Lys Ser Gly Asn Asn Asn Asp
465                 470                 475                 480


Asn Ser Asp Asp Ala Val Ala Leu Asn Val
                485                 490


<210>  148
<211>  159
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.39;
       tr|A0A1J5H6Z0|A0A1J5H6Z0_9BACT polypeptide sequence

<400>  148

Met Ala Tyr Gln Pro Ser Glu Lys Phe Met Gln Met Ala Ile Asp Lys
1               5                   10                  15


Thr Arg Glu Gly Val Leu Ser Gly Gln Thr Pro Phe Gly Ala Cys Ile
            20                  25                  30


Val Lys Asp Gly Lys Val Val Ala Cys Glu His Asn Thr Val Trp Gln
        35                  40                  45


Asp Thr Asp Ile Thr Ser His Gly Glu Val His Thr Ile Arg Ala Ala
    50                  55                  60


Cys Lys Ala Ile Gly Ser Ile Asp Leu Ser Gly Cys Ile Leu Tyr Ser
65                  70                  75                  80


Thr Cys Glu Pro Cys Pro Met Cys Phe Ser Ala Ile His Trp Ala Arg
                85                  90                  95


Ile Asp Thr Val Val Tyr Gly Ala Phe Ile Ala Asp Ala Gln Asp Ala
            100                 105                 110


Gly Phe Asn Glu Leu Thr Ile Ser Asn Glu Lys Met Lys Glu Phe Gly
        115                 120                 125


Gly Ser Pro Val Asn Phe Ile Ser Gly Phe Met Arg Asp Glu Asn Val
    130                 135                 140
```

```
Ala Leu Phe Lys Leu Trp Lys Glu Gln Gly Ala Asn Asn Val Tyr
145                 150                 155


<210>  149
<211>  168
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.40;
       tr|A0A3C2D945|A0A3C2D945_9BACT polypeptide sequence

<400>  149

Met Lys Thr Thr Glu Ile Arg Ile Ile Val His Glu Tyr Gln Asn Ile
1               5                   10                  15


Asp Glu Leu Thr Glu Asn Asp Gln Tyr Leu Leu His Glu Ala Arg Arg
            20                  25                  30


Ile Thr Glu Phe Ala Tyr Ala Pro Tyr Ser Gly Phe His Val Gly Ala
        35                  40                  45


Ala Ile Leu Leu Gly Asn Gly Met Ile Val Lys Gly Asn Asn Gln Glu
    50                  55                  60


Asn Ser Ala Tyr Pro Ser Gly Leu Cys Ala Glu Arg Val Ala Leu Phe
65                  70                  75                  80


Tyr Ala Asn Ala Asn Tyr Pro Asp Ser Glu Val Lys Thr Ile Ala Ile
                85                  90                  95


Ser Ala Ala Lys Asn Gly Ile Leu Val Asn Asp Pro Ile Lys Pro Cys
            100                 105                 110


Gly Gly Cys Arg Gln Thr Leu Ser Glu Ala Glu Val Arg Phe Gly Ser
        115                 120                 125


Pro Ile Arg Ile Ile Leu Asp Gly Gln Asp Ser Ile Leu Val Leu His
    130                 135                 140


Gly Val Glu Ser Leu Leu Pro Leu Ser Phe Ser Lys Lys Asp Leu Ala
145                 150                 155                 160
```

Ser Pro Leu Ala Ala Thr Gly Arg
165

<210> 150
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.41;
tr|A0A1I7EYS3|A0A1I7EYS3_9BURK polypeptide sequence

<400> 150

Met Lys Phe Lys Leu Asp Pro Ser Arg Pro Pro Asp Glu Asp Asp Tyr
1 5 10 15

Tyr Leu Gly Val Ala Leu Ala Val Arg Arg Lys Ala Asn Cys Thr Gly
20 25 30

Asn Arg Val Ala Ala Val Ile Val Lys Asn Lys Arg Val Ile Ala Thr
35 40 45

Gly Tyr Asn Gly Val Pro Glu Asp Met Pro Asn Cys Leu Asp Gly Gly
50 55 60

Cys Leu Arg Cys Ser Asn Pro Gly Gly Gln Phe Lys Ser Gly Thr Arg
65 70 75 80

Tyr Asp Leu Cys Ile Cys Val His Ala Glu Gln Asn Ala Leu Leu Thr
85 90 95

Ala Ala Arg Phe Gly Ile Ser Val Glu Gly Ala His Leu Tyr Thr Thr
100 105 110

Met Gln Pro Cys Phe Gly Cys Ala Lys Glu Ile Leu Gln Ala Lys Ile
115 120 125

Glu Lys Val Phe Tyr Leu His Pro Trp Val Pro Thr Asp Val Asp Pro
130 135 140

Val Met Asp Ala Ala Met Lys Ala Glu Tyr Ala Lys Ile Ile Gly Lys

145 150 155 160

Leu Lys Val Lys Lys Leu Asp Phe Asp Asp Pro Val Ala Thr Trp Ala
165 170 175

Val Thr Thr Met Arg Gln Ala Ala Leu Ala Ser Asp Lys Asn Pro Asp
180 185 190

Lys Lys Thr Pro Pro Lys Thr Ala Lys Lys Lys Val Ala Lys Lys Lys
195 200 205

Ser Arg Thr Ser Pro Arg
210

<210> 151
<211> 155
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.42;
tr|H8GQX8|H8GQX8_METAL polypeptide sequence

<400> 151

Met Asn His Glu His Phe Met Arg Arg Ala Ile Glu Leu Ala Arg Gln
1 5 10 15

Ala Pro Gln Tyr Pro Phe Gly Ala Val Ile Val Arg Arg Asp Asp Gly
20 25 30

Gln Cys Val Gly Gln Gly Phe Asn Arg Ser Asp Leu Asn Pro Thr Tyr
35 40 45

His Gly Glu Met Val Ala Ile Asn Asp Cys Ala Val Arg His Cys Ala
50 55 60

Glu Asp Trp Arg Gly Phe Asp Leu Tyr Thr Thr Ala Glu Pro Cys Ala
65 70 75 80

Met Cys Gln Gly Ala Ile Glu Trp Ala Gly Ile Gly Arg Val Phe Tyr
85 90 95

Gly Thr Ser Ile Pro Tyr Leu Gln Lys Leu Gly Trp Trp Gln Ile Asp
100 105 110

Leu Arg Ala Ala Glu Val Ser Ala Arg Ala Val Phe Arg Asp Thr Leu
115 120 125

Ile Val Gly Gly Ile Leu Glu Thr Glu Cys Asn Ala Leu Phe Ala Ala
130 135 140

Ala Arg Arg Gly Cys Phe Gly Thr Gly Ser Glu
145 150 155

<210> 152
<211> 156
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.43;
tr|A0A0S8HZN3|A0A0S8HZN3_9CHLR polypeptide sequence

<400> 152

Met Asp Glu His Asp Ile Arg Phe Leu Arg Ala Ser Phe Asp Val Ala
1 5 10 15

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

Asp Glu His Gly Arg Ile Val Met Glu Ala Glu Asn Thr Val Ile Thr
35 40 45

Ala Lys Asp Cys Thr Gly His Ala Glu Thr Asn Leu Met Arg Glu Ala
50 55 60

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

Ser Thr Glu Pro Cys Pro Met Cys Ala Gly Ala Ile Phe Trp Ser Asn
85 90 95

Val Arg Arg Val Val Tyr Gly Leu Ser Glu Glu Ser Leu Tyr Glu Ile
100 105 110

Ala Gly Arg Gly Ser Glu Glu Val Leu Phe Leu Ser Cys Arg Glu Ile
        115   120   125

Phe Glu Arg Gly Lys Lys Leu Ile Glu Val Ile Gly Pro Leu Leu Glu
    130   135   140

Asp Glu Ala Arg Glu Val His Met Gly Phe Trp Arg
145   150   155

<210> 153
<211> 193
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.44;
       tr|E3SF31|E3SF31_9CAUD polypeptide sequence

<400> 153

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

Ser Lys Cys Cys Ser Trp Lys Val Gly Ala Val Ile Glu Lys Asn Gly
            20   25   30

Arg Ile Ile Ser Thr Gly Tyr Asn Gly Ser Pro Ala Gly Gly Val Asn
        35   40   45

Cys Cys Glu His Ala Glu Glu Gln Gly Trp Leu Leu Asn Lys Pro Lys
    50   55   60

Pro Val Leu Ile Pro Gly His Lys Ser Glu Cys Val Arg Phe Ser Gln
65   70   75   80

Val Asp Arg Phe Val Leu Ala Lys Ala His Arg Glu Ala His Ser Ala
                85   90   95

Trp Ser Lys Asn Asn Glu Ile His Ala Glu Leu Asn Ala Ile Leu Phe
            100   105   110

Ala Ala Arg Met Gly Ser Ser Ile Glu Gly Ala Thr Met Tyr Val Thr
        115   120   125

```
Leu Ser Pro Cys Pro Asp Cys Ala Lys Ala Ile Ser Gln Ser Gly Ile
    130                 135                 140

Lys Lys Leu Val Tyr Cys Glu Thr Tyr Asp Lys Asn Ile Pro Gly Trp
145                 150                 155                 160

Asp Asp Ile Leu Lys Asn Ala Gly Ile Glu Val Phe Asn Val Pro Lys
                165                 170                 175

Arg Ser Leu Asp Lys Leu Asn Trp Glu Asn Ile Asn Glu Phe Cys Gly
            180                 185                 190

Glu
```

<210> 154
<211> 330
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.45; tr|F8AAC6|F8AAC6_THEID polypeptide sequence

<400> 154

```
Met Ile Arg Ala Pro Trp His Glu Tyr Phe Met Leu Leu Ala Lys Ile
1               5                   10                  15

Val Ala Leu Arg Ser Gly Cys Asn Ser Arg Pro Ser Gly Ala Val Ile
            20                  25                  30

Val Lys Asn Lys Arg Ile Leu Ala Thr Gly Tyr Asn Gly Pro Met Pro
        35                  40                  45

Gly Ala Trp His Cys Thr Asp Arg Gly Pro Gly Tyr Cys Phe Arg Arg
    50                  55                  60

Glu Lys Gly Ile Pro Asp Ile Asp Lys Tyr Asn Phe Cys Arg Ala Thr
65                  70                  75                  80

His Ala Glu Ala Asn Ala Ile Ala Gln Ala Ala Arg Phe Gly Ile Ser
```

```
                85                  90                  95

Val Glu Gly Ala Ser Leu Tyr Cys Thr Leu Ala Pro Cys Tyr Val Cys
            100                 105                 110

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

Asp Tyr Gly Ser Arg Asp Phe Glu Arg Asp Gln Phe Trp Lys Glu Ala
    130                 135                 140

Ile Lys Glu Ala Gly Leu Glu Lys Phe Glu Gln Ile Thr Val Ser Gln
145                 150                 155                 160

Glu Val Met Glu Gln Leu Gln Glu Ile Leu Pro Tyr Pro Thr Ser Lys
                165                 170                 175

Arg Arg Leu Ala Pro Thr Glu Phe Leu Asp Glu Phe Glu Asp Gly Lys
            180                 185                 190

Lys Tyr Gly Val Pro Ser Ile Glu Val Leu Phe Asn Lys Leu Asn Tyr
        195                 200                 205

Leu Thr Arg Gln Ala Leu Lys Asp Ile Thr Phe Val Ile Glu Lys Thr
    210                 215                 220

Thr Val Thr Glu Glu Pro Glu Gly Ile Ser Phe Tyr Leu Ser Gly Lys
225                 230                 235                 240

Met Val Glu Leu Ser Glu Leu Ile Asn Thr Val Lys Lys Gln Ile Asn
                245                 250                 255

Ala Asp Gln Asn Phe Tyr Phe Leu Ala Lys His Asn Ala Ile Glu Ala
            260                 265                 270

Lys Ile Glu Ile Leu Arg Glu Ala Glu Asn Ile Arg Leu Lys Ala Phe
        275                 280                 285

Leu Asn Glu Cys Pro Leu Glu Ser Phe Lys Arg Ile Ala Glu Ser Leu
    290                 295                 300
```

Asp Tyr Ile Leu Tyr Gln Val Ser Asn Ser Leu Ser Leu Pro Thr Arg
305 310 315 320

Leu Glu Leu Ser Val Asn Leu Leu Arg Ile
325 330

<210> 155
<211> 158
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.46;
tr|A0A2H4ZNK4|A0A2H4ZNK4_9EUKA polypeptide sequence

<400> 155

Met Lys Lys Gln Leu Ser Arg Lys Ile Gln Glu Glu Trp Met Ser Arg
1 5 10 15

Leu Leu Arg Asn Ala Tyr Asp Ala Gly Thr Tyr Gly Glu Val Pro Ile
20 25 30

Ala Ala Val Ile Leu Asn Glu Ser Gly Gln Cys Ile Gly Trp Gly Arg
35 40 45

Asn Cys Arg Glu Lys Asp Gln Asn Pro Leu Gly His Ala Glu Ile Ile
50 55 60

Ala Leu Arg Gln Ala Ser Tyr Leu Lys Lys Ser Trp Arg Phe Asn Glu
65 70 75 80

Cys Thr Met Leu Val Thr Leu Glu Pro Cys Pro Met Cys Ala Gly Ala
85 90 95

Leu Leu Gln Ala Arg Ile Asn His Ile Ile Tyr Gly Ala Ser Asp Tyr
100 105 110

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

Ala His His Lys Ile Glu Ile Thr Arg Gly Val Lys Ser Ile Gln Ser
130 135 140

```
Cys Gln Leu Leu Glu Thr Trp Phe Arg Arg Arg Arg Arg Val
145                 150                 155


<210>  156
<211>  176
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.47;
       tr|A0A239N5N1|A0A239N5N1_9PSED polypeptide sequence

<400>  156

Met Glu Gly Arg Ala Gly Ile Ile Pro Phe Asp Glu Gly Gly Ala Ala
1               5                   10                  15


Met Gly Pro Ala Glu Glu Asp Ser Pro Met Gln His Leu Ala Tyr Met
            20                  25                  30


Arg Glu Ala Leu Ala Leu Ala Arg Ala Asn Val Glu Ala Gly Gly Arg
        35                  40                  45


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


Ala Asn Gly Thr His Leu Asp His Asp Pro Thr Ala His Ala Glu Leu
65                  70                  75                  80


Leu Ala Leu Arg Ala Ala Gly Arg Ala Leu Gly Ser Pro Arg Leu Asp
                85                  90                  95


Gly Cys Val Val Tyr Ala Ser Gly His Pro Cys Pro Met Cys Leu Ala
            100                 105                 110


Ala Met His Leu Ser Gly Val Ser Ala Ala Tyr Tyr Ala Tyr Ser Asn
        115                 120                 125


Ala Asp Gly Glu Pro Tyr Gly Leu Ser Thr Ala Ala Val Tyr Ala Gln
    130                 135                 140


Met Ala Gln Pro Val Glu Trp Gln Ser Leu Pro Leu Gln Ala Leu Arg
```

145 150 155 160

Pro Glu Asp Glu Glu Gly Leu Tyr Gly Phe Trp Arg Glu Arg Arg Pro
165 170 175

<210> 157
<211> 199
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.48;
tr|A0A328VTR2|A0A328VTR2_9PSED polypeptide sequence

<400> 157

Met His Pro Glu His Leu Ala Leu Leu Gln Gln Ala Pro Ala Ser Thr
1 5 10 15

His Ala Asp Asp Thr Trp Ala Arg Leu Cys Cys Glu Gln Ala Leu Leu
20 25 30

Ala Val Glu Glu Gly Cys Tyr Ala Val Gly Ala Leu Leu Val Asp Gly
35 40 45

Ala Gly Glu Leu Leu Cys Ser Gly Arg Asn Gln Val Phe Ala Pro Ala
50 55 60

Tyr Ala Ser Ala Ala His Ala Glu Met Arg Val Leu Asp Gln Leu Glu
65 70 75 80

Ala Glu His Ala Gln Val Asp Arg Arg Ser Leu Thr Leu Tyr Val Ser
85 90 95

Leu Glu Pro Cys Leu Met Cys Tyr Gly Arg Ile Leu Leu Ala Gly Ile
100 105 110

Thr Arg Val Arg Tyr Leu Ala Arg Asp Arg Asp Gly Gly Phe Ala Leu
115 120 125

Arg His Gly Arg Leu Pro Pro Ala Trp Ala Asn Leu Ala Ser Gly Leu
130 135 140

```
Ser Val Val Gln Ala Lys Ala Asp Pro Tyr Trp Leu Asp Leu Ala Glu
145                 150                 155                 160


His Ala Ile Gly Arg Leu Gln Asp Arg Gln Thr Leu Arg Gln Arg Val
                165                 170                 175


Ile Arg Ala Trp Arg Gly Gln Arg Thr Leu Thr Asp Glu Phe Ser Ser
            180                 185                 190


Thr Lys Arg Thr His Ser Gly
        195


<210>  158
<211>  402
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.49;
       tr|A0A103YG48|A0A103YG48_CYNCS polypeptide sequence

<400>  158

Tyr Ile Arg Glu Leu His Ala Ser Ser Leu Arg Arg Asp Glu His Glu
1               5                   10                  15


Ile Gln Asn Pro Lys Ile Leu Val Ile Val Asp Arg Leu Ser Ser Pro
            20                  25                  30


Ser Leu His Val Ser Leu Ser Leu Ser Leu Ser Leu Val Ile Phe Pro
        35                  40                  45


Pro Phe Ile Pro Leu Asn Gln Thr Pro Thr His Met Glu Asn Ala Lys
    50                  55                  60


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


Gly His Gln Glu Val Val Gln Asp Arg Asp His Lys Phe Leu Thr Arg
                85                  90                  95


Ala Val Glu Glu Ala Tyr Lys Gly Val Glu Cys Gly Asp Gly Gly Pro
            100                 105                 110
```

```
Phe Gly Ala Val Val Val His Lys Asp Glu Val Val Ala Ser Cys His
        115                 120                 125

Asn Met Val Leu Lys His Thr Asp Pro Thr Ala His Ala Glu Val Thr
    130                 135                 140

Ala Ile Arg Glu Ala Cys Lys Lys Leu Asn Lys Ile Glu Leu Ser Asp
145                 150                 155                 160

Cys Glu Ile Tyr Ala Ser Cys Glu Pro Cys Pro Met Cys Phe Gly Ala
                165                 170                 175

Ile His Leu Ser Arg Ile Lys Arg Leu Ile Tyr Gly Ala Lys Ala Glu
            180                 185                 190

Ala Ala Ile Ala Ile Gly Phe Asp Asp Phe Ile Ala Asp Ala Leu Arg
        195                 200                 205

Gly Thr Gly Phe Tyr Gln Lys Ala His Leu Glu Ile Lys Gln Ala Asp
    210                 215                 220

Gly Asn Gly Ala Met Ile Ala Glu Gln Val Phe Glu Lys Thr Lys Ala
225                 230                 235                 240

Lys Phe Ala Ile Asp His Lys Phe Leu Thr Arg Ala Val Glu Glu Ala
                245                 250                 255

Tyr Lys Gly Val Glu Cys Gly Asp Gly Arg Pro Phe Gly Ala Leu Val
            260                 265                 270

Val His Lys Asp Glu Val Val Val Ser Cys His Asn Met Val Leu Asn
        275                 280                 285

Tyr Thr Asp Pro Thr Ala His Ala Glu Ile Thr Ala Ile Arg Glu Ala
    290                 295                 300

Cys Lys Lys Leu Asn Arg Ile Glu Leu Ser Asp Cys Glu Met Tyr Ser
305                 310                 315                 320

Ser Cys Glu Pro Cys Pro Met Cys Phe Gly Ala Ile Gln Ile Ser Arg
```

```
                325                 330                 335

Ile Lys Arg Leu Val Tyr Gly Ala Lys Ala Glu Ala Ser Ile Ala Ser
            340                 345                 350

Gly Ile Pro Ile Gly Asp Phe Ile Ser Asp Ala Leu Lys Gly Thr Gly
        355                 360                 365

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

Ala Met Ile Ala Glu Gln Val Phe Glu Arg Thr Lys Ala Met Phe Pro
385                 390                 395                 400

Lys Arg
```

<210> 159
<211> 195
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.50; tr|W5M1M8|W5M1M8_LEPOC polypeptide sequence

<400> 159

```
Asn Ser Ser Thr Arg Glu Ser Arg Val Met Ala Gln Met Glu Ile Asn
1               5                   10                  15

Gly Gly Ala Ser Pro Pro Lys Lys Pro Gly Lys Gly Gln Ser Ala Ala
            20                  25                  30

Asp Gln Asp Met Ile Thr Gly Leu Ile Asn Lys Ala Leu Gln Ala Lys
        35                  40                  45

Glu Phe Ala Tyr Cys Pro Tyr Ser Asn Phe Arg Val Gly Ala Ala Leu
    50                  55                  60

Met Thr Asn Asp Gly Arg Val Phe Thr Gly Cys Asn Val Glu Asn Ala
65                  70                  75                  80
```

```
Cys Tyr Asn Leu Gly Val Cys Ala Glu Arg Thr Ala Ile Leu Lys Ala
                85                  90                  95

Val Ser Glu Gly Tyr Glu Ser Phe Arg Ala Ile Ala Val Ser Ser Asp
            100                 105                 110

Leu Gln Asp Gln Phe Ile Ser Pro Cys Gly Ala Cys Arg Gln Val Met
        115                 120                 125

Arg Glu Phe Gly Thr Gly Trp Asp Val Phe Leu Thr Lys Val Asp Gly
    130                 135                 140

Ser Tyr Val Arg Met Thr Val Asp Glu Leu Leu Pro Met Ser Phe Gly
145                 150                 155                 160

Pro Asp Asp Leu Lys Lys Lys Lys Val Phe Ser Leu Gln Asn Gly His
                165                 170                 175

Glu Val Ser Thr Gln Phe Tyr Thr His Ser Pro Cys Glu Ala Gly Glu
            180                 185                 190

Asn Asn Asn
        195
```

<210> 160
<211> 134
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.51;
       tr|A0A3N5YPZ2|A0A3N5YPZ2_9ALTE polypeptide sequence

<400> 160

```
Met Ser Asn Ser Glu Thr Glu His Ile Gln Ala Leu Val Asp Ala Ala
1               5                   10                  15

Gln Ala Ala Gln Lys Gln Ser Tyr Ser Pro Tyr Ser Ser Phe Gln Val
            20                  25                  30

Gly Ala Ala Ile Phe Ala Asp Asp Gly Asn Thr Tyr Ser Gly Cys Asn
        35                  40                  45
```

```
Ile Glu Asn Val Ala Tyr Pro Leu Gly Gln Cys Ala Glu Ala Thr Ala
    50                  55                  60

Ile Gly Met Met Ile Met Gln Gly Ala Lys Arg Ile Glu Asp Ile Met
65                  70                  75                  80

Ile Ala Ser Pro Asn Asp Gln Val Cys Pro Pro Cys Gly Gly Cys Arg
                85                  90                  95

Gln Lys Ile Ser Glu Phe Gly Thr Ala Glu Thr Lys Ile His Met Val
            100                 105                 110

Thr Arg Ser Gly Glu Val Ser Thr Val Thr Leu Gly Glu Leu Leu Pro
        115                 120                 125

Leu Ala Phe Asp Ser Leu
    130
```

<210> 161
<211> 162
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.52;
       tr|A0A2A9NC86|A0A2A9NC86_9AGAR polypeptide sequence

<400> 161

```
Met Thr Asn Ser Thr Leu Ser Asn Glu Asp Arg Thr Arg Leu Ile Gln
1               5                   10                  15

Gly Ala Phe Gln Ala Arg Lys Lys Thr Tyr Ser Pro Tyr Ser Asn Phe
            20                  25                  30

Pro Val Gly Ala Ala Leu Leu Thr Thr Asp Gly Arg Ile Ile Glu Gly
        35                  40                  45

Ala Asn Ile Glu Asn Ala Ser Tyr Gly Gly Thr Ile Cys Ala Glu Arg
    50                  55                  60

Thr Ala Ile Val Lys Ala Val Ser Asp Gly Tyr Arg His Phe Ala Gly
65                  70                  75                  80
```

```
Ile Ala Val Thr Thr Lys Met Pro Thr Arg Val Ser Pro Cys Gly Ile
                85                  90                  95

Cys Arg Gln Val Leu Arg Glu Phe Cys Ser Leu Asp Met Pro Val Leu
            100                 105                 110

Leu Val Pro Gly Asp Tyr Pro Gln Arg Asn Pro Val Asp Asp Asp Gly
        115                 120                 125

Ala Asp Lys Pro Gly Val Ile Thr Glu Gly Gly Val Arg Glu Thr Thr
    130                 135                 140

Leu Gly Ala Leu Leu Pro Asp Ser Phe Gly Pro Glu Asn Leu Pro Pro
145                 150                 155                 160

Arg Ala
```

<210> 162
<211> 163
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.53;
tr|A0A2D6RD43|A0A2D6RD43_9GAMM polypeptide sequence

<400> 162

```
Met Asn Ile Glu Asn Leu Ile Thr Glu Asn Asp Glu Thr Leu Ile Arg
1               5                   10                  15

Arg Cys Ile Glu Leu Ala Gly Glu Ser Val Lys Asn Gly Asp Lys Pro
            20                  25                  30

Phe Gly Ala Leu Leu Ala Lys Asp Gly Asn Ile Ile Phe Glu Ser Ser
        35                  40                  45

Asn Asn Ala Lys Thr Lys Val Pro Tyr His Ala Glu Ile Leu Thr Leu
    50                  55                  60

Met Asp Ala Gln Asp Lys Leu Asn Thr Thr Asp Leu Ser Asp Tyr Ala
```

65 70 75 80

Leu Tyr Ser Asn Cys Glu Pro Cys Pro Met Cys Ser Phe Met Ile Arg
85 90 95

Glu Tyr Lys Leu Asp Lys Val Val Phe Ser Val His Ser Pro Tyr Met
100 105 110

Gly Gly Gln Ser Arg Trp Asn Ile Leu Glu Asp Asp Val Leu Thr Arg
115 120 125

Phe Lys Pro Tyr Phe Ser Lys Pro Pro Asn Val Val Gly Gly Val Leu
130 135 140

Glu Ser Glu Gly Lys Arg Ile Phe Asp Lys Val Gly Leu Trp Met Phe
145 150 155 160

Gly Lys Glu

<210> 163
<211> 204
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.54;
tr|A0A0H3AVL6|A0A0H3AVL6_BRUO2 polypeptide sequence

<400> 163

Met His Ala Lys Gly Tyr Ser Gln Gln Glu Arg Arg Ile Ile Pro Phe
1 5 10 15

Ala Asn Arg Phe Arg Phe Arg Glu Leu Cys Ser Asn Lys Ser Leu His
20 25 30

Gly Leu Arg Ala Lys Phe Pro Glu Gln Tyr Thr Lys Trp Asp Pro Met
35 40 45

Arg Lys Ala Ala Ser Ile Thr Lys Ala Asn Ser Ala Thr Pro Met Asp
50 55 60

```
Ile Ala Leu Glu Glu Ala His Ala Ala Gly Glu Arg Gly Glu Val Pro
65                  70                  75                  80


Ile Gly Ala Val Ile Val Arg Asp Gly Glu Ile Ile Ala Arg Ala Gly
                85                  90                  95


Asn Arg Thr Arg Glu Phe Asn Asp Val Thr Ala His Ala Glu Ile Leu
            100                 105                 110


Thr Ile Arg Gln Ala Gly Glu Met Leu Gly Ser Glu Arg Leu Ile Asp
        115                 120                 125


Cys Asp Leu Tyr Val Thr Leu Glu Pro Cys Ala Met Cys Ala Ala Ala
    130                 135                 140


Ile Ser Phe Ala Arg Ile Arg Arg Leu Tyr Tyr Gly Ala Ser Asp Pro
145                 150                 155                 160


Lys Gly Gly Gly Ile Glu His Gly Gly Arg Phe Tyr Thr Gln Pro Thr
                165                 170                 175


Cys His His Ala Pro Glu Ile Tyr Pro Gly Phe Cys Glu Ala Asp Ala
            180                 185                 190


Arg Lys Ile Leu Lys Asp Phe Phe Arg Glu Lys Arg
        195                 200


<210>  164
<211>  165
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.55;
       tr|A0A242H531|A0A242H531_9ENTE polypeptide sequence

<400>  164

Met Phe Ile Val Lys Asn Asn Ile Glu Val Ile Gln Gln Gln Ala Glu
1               5                   10                  15


Leu Asp Ala Lys Phe Met Lys Gln Ala Leu Lys Leu Ala Lys Asp Ala
            20                  25                  30
```

```
Ser Asn Asn Gly Asn Glu Pro Phe Gly Ala Val Leu Val Lys Asn Asp
        35                  40                  45

Lys Val Ile Leu Thr Gly Glu Asn Gln Ile His Thr Glu Ser Asp Pro
    50                  55                  60

Thr Tyr His Ala Glu Leu Gly Ile Ile Arg Asp Phe Cys Thr Ser Gln
65                  70                  75                  80

Lys Ile Thr Asp Leu Ser Glu Tyr Thr Leu Tyr Thr Ser Cys Glu Pro
                85                  90                  95

Cys Cys Met Cys Ala Gly Ala Met Val Trp Ser Asn Leu Asp Arg Met
            100                 105                 110

Val Tyr Gly Leu Gly His Asp Glu Leu Ala Glu Ile Ala Gly Phe Asn
        115                 120                 125

Ile Met Ile Gly Ser Glu Glu Ile Phe Ser Lys Ser Pro Asn Arg Pro
    130                 135                 140

Glu Val Ala Lys Gly Val Leu Lys Glu Ala Ala Val Pro Val Tyr Val
145                 150                 155                 160

Asp Tyr Phe Gln Arg
                165
```

<210> 165
<211> 181
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.56;
       tr|A0A2R6XZE2|A0A2R6XZE2_9BACL polypeptide sequence

<400> 165

```
Met Ser Gly Arg Ile Ser Trp His Glu Tyr Phe Met Ala Gln Ala Lys
1               5                   10                  15

Leu Ile Ala Leu Arg Ala Thr Cys Thr Arg Leu Met Val Gly Ala Val
            20                  25                  30
```

```
Ile Val Arg Asp Arg Arg Val Ile Ala Gly Gly Tyr Asn Gly Ser Ile
        35                  40                  45

Ala Gly Asp Glu His Cys Ile Asp Val Gly Cys Lys Val Arg Asp Gly
    50                  55                  60

His Cys Ile Arg Thr Ile His Ala Glu Gln Asn Ala Leu Met Gln Cys
65                  70                  75                  80

Ala Lys Phe Gly Val Ser Thr Asp Gly Ala Glu Leu Tyr Val Thr His
                85                  90                  95

Phe Pro Cys Leu Asn Cys Thr Lys Leu Leu Ile Gln Ala Gly Ile Arg
            100                 105                 110

His Ile Tyr Tyr Glu Val Pro Tyr Arg Val Asp Pro Tyr Ala Ile Glu
        115                 120                 125

Leu Leu Glu Lys Ala Gly Val Gly Thr Thr Gln Ile Thr Val Asp Leu
    130                 135                 140

Asn Ala Tyr Val Gln Val Met Ser Lys Val Ser Thr Asp Pro Ala Leu
145                 150                 155                 160

Thr Tyr Val Pro Glu Ser Lys Ala Gln Lys Asp Glu Tyr Gly Gln Ser
                165                 170                 175

Val Gly Lys Ile Val
            180
```

<210> 166
<211> 173
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.57;
tr|A0A139SHT6|A0A139SHT6_9BACT polypeptide sequence

<400> 166

```
Met Ser Glu Ala Asn Ala Ser Ser Glu Ser Leu Pro Ser Arg Asn Ser
```

```
1               5                   10                  15

Pro Val Glu Leu Ile Ala Glu Ala Ala Gly Lys Phe Gly Arg Arg Pro
            20                  25                  30

Thr Trp Asp Glu Tyr Phe Met Ala Thr Ala Val Leu Ile Ser Thr Arg
        35                  40                  45

Ser Ser Cys Glu Arg Leu Asn Val Gly Cys Val Ile Val Thr Ala Gly
    50                  55                  60

Glu Ser His Lys Asn Arg Ile Val Ala Ala Gly Tyr Asn Gly His Leu
65                  70                  75                  80

Pro Gly Ser Pro His Thr Ser Arg Met Arg Asp Gly His Glu Gln Ala
                85                  90                  95

Thr Val His Ala Glu Gln Asn Ala Ile Ser Asp Ala Ala Arg Arg Gly
            100                 105                 110

Ser Ser Val Glu Gly Cys Thr Ala Tyr Val Thr His Tyr Pro Cys Ile
        115                 120                 125

Asn Cys Ala Lys Ile Leu Ala Ser Ala Gly Ile Ala Lys Ile Cys Tyr
    130                 135                 140

Arg Leu Asp Tyr His Asn Asp Pro Leu Val Lys Pro Met Leu Ala Glu
145                 150                 155                 160

Ala Gly Ile Glu Ile Val Gln Leu Gly Glu Ala Ala Ser
                165                 170
```

<210> 167
<211> 170
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.58; tr|A0A261DBH2|A0A261DBH2_9RICK polypeptide sequence

<400> 167

```
Met Val Met Lys Lys Lys Leu Ile Thr Val Lys Arg Ser Thr Glu Phe
1               5                   10                  15

Asn Asn Phe Phe Met Glu Glu Ala Leu Lys Gln Ala Gln Phe Ala Leu
            20                  25                  30

Asp Lys Asn Glu Ile Pro Val Gly Ala Ile Ile Val Asn Arg Ile Thr
        35                  40                  45

Asn Lys Val Ile Ala Lys Ala His Asn Ile Val Glu Gln Thr Lys Asn
    50                  55                  60

Pro Val Leu His Ala Glu Ile Val Ala Ile Asn Gln Ser Cys Gln Ile
65                  70                  75                  80

Leu Ser Ser Lys Asn Leu Ser Asp Cys Asp Met Tyr Val Thr Leu Glu
                85                  90                  95

Pro Cys Val Met Cys Ser Gly Ala Ile Ser Phe Ala Arg Ile Gly Arg
            100                 105                 110

Leu Phe Tyr Ala Ala Asn Asp Pro Lys Gln Gly Ala Ile Glu Asn Gly
        115                 120                 125

Gly Arg Phe Phe Asn Ser Lys Ser Cys Phe Tyr Arg Pro Glu Ile Tyr
    130                 135                 140

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

Tyr Asn Val Arg Tyr Gln Lys Cys Asn Pro
                165                 170
```

<210> 168
<211> 146
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.59;
tr|A0A2N0XZK6|A0A2N0XZK6_9VIBR polypeptide sequence

<400> 168

```
Met Thr Asp Asn Ser Leu His Glu Ser Tyr Met Arg Gln Ala Phe Glu
1               5                  10                  15

Leu Ser Lys Ser Ala Leu Pro Gly Cys Arg Pro Asn Pro Pro Val Gly
            20                  25                  30

Cys Val Phe Val Lys Asp Gly Glu Val Val Ser Ser Gly Phe Ser Gln
        35                  40                  45

Pro Pro Gly Asn His His Ala Glu Ala Gly Ala Ile Ala Ala Tyr Thr
    50                  55                  60

Gly Ser Tyr Asp Gly Leu Val Ala Tyr Val Thr Leu Glu Pro Cys Ser
65                  70                  75                  80

Phe Gln Gly Arg Thr Pro Ser Cys Ala Lys Ala Leu Val Arg Val Arg
                85                  90                  95

Pro Glu Lys Val Tyr Val Ala Ile Leu Asp Pro Asp Thr Arg Asn Ser
            100                 105                 110

Gly Ala Gly Ile Lys Ile Leu Glu Asp Ala Gly Ile Asp Val Glu Val
        115                 120                 125

Gly Leu Leu Gly Glu Glu Val Ala Ser Phe Leu Asn Pro Tyr Leu Ile
    130                 135                 140

Arg Asn
145
```

<210> 169
<211> 159
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.60;
       tr|A0A1V5R0F9|A0A1V5R0F9_9BACT polypeptide sequence

<400> 169

```
Met Thr Lys Lys Glu Thr Thr Lys Leu His Ala Leu Asp Asp Phe Cys
1               5                  10                  15
```

```
Met Lys Lys Ala Leu Leu Leu Ala Lys Arg Ala Phe Arg Ala Asp Glu
            20                  25                  30

Val Pro Val Gly Ala Leu Val Val Asp Ser Ser Asn Lys Val Ile Gly
        35                  40                  45

Arg Gly Tyr Asn Gln Val Glu Lys Arg Lys Ser Gln Arg Ala His Ala
    50                  55                  60

Glu Gln Leu Ala Ile Glu Gln Ala Cys Lys Lys Ile Gly Asp Trp Arg
65                  70                  75                  80

Leu Glu Gly Cys Thr Leu Tyr Val Thr Leu Glu Pro Cys Thr Met Cys
                85                  90                  95

Met Gly Leu Ile Lys Leu Ser Arg Ile Glu Arg Val Val Phe Gly Ala
            100                 105                 110

Ala Ser Pro Leu Phe Gly Tyr Gln Leu Asp Lys Asn Arg Lys Ser Gln
        115                 120                 125

Leu Tyr Lys Lys Gly Val Ile Lys Ile Arg Lys Gly Val Gly Lys Ala
    130                 135                 140

Thr Ala Ala Ala Leu Leu Lys Asp Phe Phe Lys Asn Lys Arg Met
145                 150                 155
```

<210> 170
<211> 167
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.61;
       tr|A0A2W0H8Y3|A0A2W0H8Y3_9BACI polypeptide sequence

<400> 170

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

Leu Gln Glu Ala Lys Glu Ala Gly Gln Arg Gly Asp Leu Pro Ile Gly
```

```
            20                  25                  30

Ala Val Ile Val His Asn Gly Arg Ile Ile Ala Arg Gly Ser Asn Met
        35                  40                  45

Arg Lys Thr Ala Gly Ile Lys Ile Ser His Ala Glu Asn Asn Ala Met
    50                  55                  60

His Asn Cys Ala Pro Tyr Leu Met Lys His Ala Ser Glu Cys Val Ile
65                  70                  75                  80

Tyr Thr Thr Leu Glu Pro Cys Ile Met Cys Leu Thr Thr Leu Val Met
                85                  90                  95

Ala Asn Ile Asp Ser Ile Val Phe Ala Ala Asp Asp Lys Tyr Met Asn
            100                 105                 110

Met Lys Pro Phe Ile Asp Ala Asn Ser Tyr Ile Arg Asp Arg Ile His
        115                 120                 125

Gln Tyr Lys Gly Gly Val Cys Arg Gly Glu Ser Glu Ala Leu Leu Arg
    130                 135                 140

Lys Tyr Ser Pro Tyr Ala Ala Glu Leu Ala Leu Asn Gly Thr His Pro
145                 150                 155                 160

His His Arg Lys Gly Gly Ala
                165
```

<210> 171
<211> 165
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.62;
tr|A0A261BDB7|A0A261BDB7_CAERE polypeptide sequence

<400> 171

```
Leu Tyr Lys Leu Tyr Ile Phe Arg Met Thr Thr Thr Lys Ala Asn Leu
1               5                   10                  15
```

```
Thr Gln Phe Glu Gln Glu Leu Val Asp Lys Ala Val Gly Ala Met Glu
            20                  25                  30

Lys Ala Tyr Cys Lys Tyr Ser Gly Phe Lys Val Gly Ala Ala Leu Val
        35                  40                  45

Cys Glu Asp Gly Glu Ile Ile Ile Gly Ala Asn His Glu Asn Ala Ser
    50                  55                  60

Tyr Gly Ala Thr Ile Cys Ala Glu Arg Ser Ala Met Val Thr Ala Leu
65                  70                  75                  80

Thr Lys Gly His Arg Lys Phe Lys Leu Leu Ala Val Ala Thr Glu Leu
                85                  90                  95

Glu Ala Pro Cys Ser Pro Cys Gly Ile Cys Arg Gln Tyr Leu Ile Glu
            100                 105                 110

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

Ile Glu Thr Thr Thr Tyr Gly Leu Leu Pro Tyr Ala Phe Thr Pro Lys
    130                 135                 140

Ser Leu Asp Asp His Glu Lys Glu Ala Glu Glu Arg Asn His Gln Glu
145                 150                 155                 160

Gly Glu Lys Lys His
                165
```

<210> 172
<211> 162
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.63; tr|A0A2E1PHI6|A0A2E1PHI6_9GAMM polypeptide sequence

<400> 172

```
Met Lys Glu Leu Leu Ile His Ser Trp Leu Met Leu Asn Ser Asn Ser
1               5                   10                  15
```

```
Lys Leu Ile Met Glu Arg Val Ile Glu Leu Ser Glu Ile Asn Leu Lys
            20                  25                  30

Asn Gly Lys Ile Pro Ile Ala Ala Val Ile Val Asp Lys Lys Asn Tyr
        35                  40                  45

Glu Ile Ile Ser Glu Ser Gln Asn Glu Asp Ser Pro Ile Gly His Ala
    50                  55                  60

Glu Leu Leu Ala Ile Thr Lys Ala Leu Lys Lys Leu Asn Thr Asn Arg
65                  70                  75                  80

Leu Asp Ser Thr Asn Leu Phe Val Thr Ile Glu Pro Cys Pro Met Cys
                85                  90                  95

Ala Tyr Ala Ile Ser Lys Cys His Ile Asn Arg Leu Tyr Phe Gly Ser
            100                 105                 110

Glu Asp Glu Lys Gly Gly Gly Val Ile Asn Gly Pro Arg Ile Phe Glu
        115                 120                 125

Ser His Asn Leu Lys Lys Ile Asp Tyr Val Ser His Cys Tyr His Glu
    130                 135                 140

Lys Thr Thr Gln Leu Met Gln Ser Phe Phe Gln Leu Lys Arg Asn Gln
145                 150                 155                 160

Gln Leu
```

<210> 173
<211> 133
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.64; tr|A0A378LUA7|A0A378LUA7_9GAMM polypeptide sequence

<400> 173

```
Met Asp Thr Ile Ile Lys Lys Met Ile Ser Asn Ala His Asn Thr Leu
1               5                   10                  15
```

```
Ala His Ser Tyr Ser Pro Tyr Ser Lys Phe Ser Val Ala Ser Cys Ile
            20                  25                  30

Cys Thr Asp Lys Asp Asn Phe Tyr Thr Gly Val Asn Val Glu Asn Ser
        35                  40                  45

Ala Tyr Gly Leu Ala Ile Cys Ala Glu Thr Ser Ala Ile Ser Ala Met
    50                  55                  60

Val Thr Ala Gly Glu Lys Arg Ile Lys Ser Met Val Val Met Ala Gly
65                  70                  75                  80

Thr Asn Ile Leu Cys Ser Pro Cys Gly Ala Cys Arg Gln Arg Ile Tyr
                85                  90                  95

Glu Phe Ser Thr Pro Asp Thr Leu Ile His Leu Cys Asp Lys Asn Ser
            100                 105                 110

Ile Leu Arg Thr Phe Lys Ile Asn Glu Leu Leu Pro Glu Ala Phe Lys
        115                 120                 125

Phe Asp Phe Asn Pro
    130
```

<210> 174
<211> 370
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.65; tr|A0A139HQ78|A0A139HQ78_9PEZI polypeptide sequence

<400> 174

```
Met Ala Asp Ser Leu Lys Ser Lys Pro Gly His Ala Arg His Asp Thr
1               5                   10                  15

Ala Leu Ile His Gly Leu Ser Gln Ser Asp Val Gln Lys Leu Ser Glu
            20                  25                  30

Ser Cys Val Asp Ala Lys Ser Lys Ala Tyr Cys Pro Tyr Ser His Phe
```

```
        35                  40                  45

Arg Val Gly Cys Ala Val Leu Leu Ala Asn Gly Asp Val Val Gln Gly
    50                  55                  60

Ala Asn Val Glu Asn Ala Ala Tyr Pro Val Gly Thr Cys Ala Glu Arg
65                  70                  75                  80

Val Ala Leu Gly Thr Ala Val Gly Ala Lys Lys Gly Asp Phe Arg Ala
                85                  90                  95

Leu Ala Val Ser Thr Asp Ile Ser Pro Pro Ala Ser Pro Cys Gly Met
            100                 105                 110

Cys Arg Gln Phe Ile Arg Glu Phe Cys Glu Leu Asn Thr Pro Ile Leu
        115                 120                 125

Met Tyr Asp Lys Asp Gly Lys Ser Val Val Met Thr Leu Glu Gln Leu
    130                 135                 140

Leu Pro Met Ser Phe Gly Pro Asp Lys Leu Leu Pro Pro Gly Gln Leu
145                 150                 155                 160

Glu Asn Gly Leu Met Gln Thr Gln Thr Gln Ser Ser Phe Val Thr Arg
                165                 170                 175

Ala Phe Ser Thr Thr Ser Ser Arg Arg Gln Asp Asp Thr Pro Gln Val
            180                 185                 190

Pro Gln Ser His Tyr Asp Phe Phe Pro Gln Thr Phe Pro Gln Gly Pro
        195                 200                 205

Pro Pro Lys Thr Ser Phe Ser Pro Asp Leu Lys Gln Leu Arg Lys Glu
    210                 215                 220

Phe Leu Gln Leu Gln Ala Lys Ala His Pro Asp Leu Ala Pro Gln Asp
225                 230                 235                 240

Gln Lys Arg Arg Ala Glu Ala Leu Ser Met Arg Ile Asn Glu Ala Tyr
                245                 250                 255
```

```
Lys Thr Leu Gln Ser Pro Leu Arg Arg Ala Gln Tyr Leu Leu Ser Gln
            260                 265                 270


Gln Gly Ile Asp Val Glu Asp Glu Thr Ala Lys Leu Asp Asp Ser Ser
        275                 280                 285


Leu Leu Met Glu Val Met Glu Ala Arg Glu Ala Val Glu Glu Val Glu
    290                 295                 300


Asp Glu Glu Gln Leu Asn Glu Ile Arg Ala Glu Asn Asn Gly Arg Ile
305                 310                 315                 320


Glu Glu Ser Val Arg Val Leu Glu Asp Ala Phe Arg Asp Asn Glu Phe
                325                 330                 335


Glu Lys Ala Ala Gln Glu Ala Ile Arg Leu Arg Tyr Trp Val Asn Ile
            340                 345                 350


Glu Glu Ser Ile Gln Gly Trp Glu Lys Gly Asn Gly Gly Gly Ile Leu
        355                 360                 365


His His
    370


<210>  175
<211>  165
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.66;
       tr|A0A2A9FXV0|A0A2A9FXV0_9VIBR polypeptide sequence

<400>  175

Met Cys Asn Leu Lys Glu Asn Lys Asp Met Asp Lys Tyr Phe His Phe
1               5                   10                  15


Ala Cys Asp Ala Thr Ile Glu Gly Met Arg Glu Gly Thr Gly Gly Pro
            20                  25                  30


Phe Gly Ala Thr Leu Thr Arg Asn Gly Glu Val Val Cys Ser Val Ala
        35                  40                  45
```

```
Asn Thr Val Leu Lys Asp Met Asp Ile Ser Gly His Ala Glu Met Val
    50                  55                  60

Ala Val Arg Glu Ala Cys Lys Lys Leu Asp Thr Leu Asp Leu Ser Asp
65                  70                  75                  80

Cys Val Met Tyr Ala Thr Cys Glu Pro Cys Pro Met Cys Val Ser Val
                85                  90                  95

Met Leu Trp Ala Gly Ile Lys Thr Cys Tyr Tyr Ala Ser Thr His Leu
            100                 105                 110

Asp Ala Ala Lys His Gly Phe Ser Asp Gln Gln Leu Arg Asp Tyr Leu
        115                 120                 125

Asp Gly Ser Asp Thr Ser Thr Leu Asn Met Val His Ile Glu Asp Asn
    130                 135                 140

Arg Asp Asp Cys Ala Lys Ile Trp Thr Glu Phe Arg His Leu Asn Glu
145                 150                 155                 160

Thr Lys Asn Asp Gly
                165
```

<210> 176
<211> 509
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.67; tr|A0A1A8AG96|A0A1A8AG96_NOTFU polypeptide sequence

<400> 176

```
Met Glu His Ser Asp Arg Trp Ser Arg Ala Glu Pro Gly Leu Ser Thr
1               5                   10                  15

Ser Ser Arg Glu Thr Arg Asp Gly Ser Thr Gln Thr Asp Cys Lys Leu
            20                  25                  30

Gln Gly His Gly Pro Arg Leu Ser Lys Val Asn Leu Phe Thr Leu Leu
```

```
        35                  40                  45

Ser Leu Trp Met Glu Leu Phe Pro Gln Glu Gln Asp Glu Glu Asn Gly
    50                  55                  60

Gln Ser Gln Ile Arg Arg Ser Gly Leu Val Val Val Arg Glu Gly Lys
65                  70                  75                  80

Val Val Gly Leu His Cys Ser Gly Ala Asp Leu His Ala Gly Gln Ala
                85                  90                  95

Ala Ile Leu Gln His Gly Ala Ser Leu Ala Asn Cys Gln Leu Phe Phe
            100                 105                 110

Ser Arg Arg Pro Cys Ala Thr Cys Leu Lys Met Ile Ile Asn Ala Gly
        115                 120                 125

Val Arg Gln Ile Thr Phe Trp Pro Gly Asp Pro Glu Ile Ser Met Leu
    130                 135                 140

Thr Ser Asn Gln Thr His Ser Gln Arg Thr Ser Gln Ser Ile Thr Glu
145                 150                 155                 160

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

Pro Gln Ile Cys Val Leu Met Gln Pro Leu Ala Pro Gly Val Leu Gln
            180                 185                 190

Phe Val Asp Glu Thr Ser Arg Arg Ser Asp Phe Met Glu Arg Met Met
        195                 200                 205

Asp Asp Asp Pro Glu Leu Asp Ser Glu Lys Leu Phe Asn Ser Asp Arg
    210                 215                 220

Leu Arg His Leu Lys Asp Phe Cys Arg His Phe Leu Ile Gln Thr Asp
225                 230                 235                 240

Gln Arg His Lys Asp Ile Leu Ser Gln Met Gly Leu Lys Asn Phe Cys
                245                 250                 255
```

```
Val Glu Pro Tyr Phe Ser Asn Leu Arg Ser Asn Met Thr Glu Leu Val
            260                 265                 270

Glu Val Leu Ala Ala Val Ala Ala Gly Met Pro Gln Gln His Tyr Gly
        275                 280                 285

Phe Tyr Arg Glu Glu Ser Leu Ser Leu Asp Pro His Pro Val Asp Val
    290                 295                 300

Ser Gln Ala Val Ala Arg His Cys Ile Val Gln Ala Arg Leu Leu Ser
305                 310                 315                 320

Tyr Arg Thr Glu Asp Pro Lys Val Gly Val Gly Ala Val Ile Trp Ala
                325                 330                 335

Lys Gly Gln Ser Ala Cys Cys Cys Gly Thr Gly Arg Leu Tyr Leu Ile
            340                 345                 350

Gly Cys Gly Tyr Asn Ala Tyr Pro Ala Gly Ser Lys Tyr Ala Glu Tyr
        355                 360                 365

Pro Gln Met Asp Asn Lys Gln Glu Asp Arg Glu Arg Arg Lys Tyr Arg
    370                 375                 380

Tyr Ile Val His Ala Glu Gln Asn Ala Leu Thr Phe Arg Thr Arg Asp
385                 390                 395                 400

Ile Lys Pro Asp Glu Cys Ser Met Leu Phe Val Thr Lys Cys Pro Cys
                405                 410                 415

Asp Glu Cys Ile Pro Leu Ile Arg Gly Ala Gly Val Lys His Ile Tyr
            420                 425                 430

Thr Ser Asp Gln Asp Arg Asp Lys Asp Lys Gly Asp Ile Ser Tyr Leu
        435                 440                 445

Arg Phe Gly Ser Leu Lys Gly Val Cys Lys Phe Ile Trp Gln Arg Ser
    450                 455                 460

Pro Pro Val Ser Ser Ala Ser Ser Leu His Leu Thr Asn Gly Cys Val
```

```
465                 470                 475                 480

Gly Lys His Val Arg Gln Ala Glu Gln Gln Ile Tyr Lys Asn Lys Lys
                485                 490                 495

Leu Cys Thr Lys Gly Ser Ser Gly Ser Ser Asp Ile Cys
            500                 505
```

<210> 177
<211> 148
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.68; tr|A0A3E2VN88|A0A3E2VN88_9FIRM polypeptide sequence

<400> 177

```
Met Glu Lys Glu Ile Thr Asn Met Asp Lys Gln Lys Leu Ile Gln Met
1               5                   10                  15

Ala Val Asp Gly Leu Gly Arg Ser Tyr Ala Pro Tyr Ser His Phe His
            20                  25                  30

Val Ser Ala Ala Leu Leu Cys Ala Asp Gly Thr Val Tyr Thr Gly Asn
        35                  40                  45

Asn Ile Glu Asn Ala Ala Tyr Thr Pro Ser Val Cys Ala Glu Arg Cys
    50                  55                  60

Ala Ile Phe Lys Ala Val Gly Asp Gly Arg Arg Glu Phe Glu Ala Ile
65                  70                  75                  80

Ala Val Cys Gly Gly Pro Asp Gly Val Ile Glu Asp Tyr Cys Pro Pro
                85                  90                  95

Cys Gly Val Cys Arg Gln Val Met Arg Glu Phe Cys Asp Pro Ser Ser
            100                 105                 110

Phe Arg Val Leu Val Ala Lys Thr Ala Glu Asp Tyr Arg Glu Tyr Thr
        115                 120                 125
```

Leu Glu Gln Leu Leu Pro Asp Gly Phe Gly Pro Asp His Leu Thr Gly
130 135 140

Ser Gly Glu Arg
145

<210> 178
<211> 221
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.69;
tr|A0A2D5ZRJ2|A0A2D5ZRJ2_9BACT polypeptide sequence

<400> 178

Met Ala Arg Pro Val His Leu His Thr Gly Glu Arg Arg Thr Glu Glu
1 5 10 15

Gly Ala Thr Glu Ser Arg Ala Val Ala Ala Val Ala Thr Ala Ile Thr
20 25 30

Arg Ala Pro Arg Ala Pro Pro Arg Pro Ala Thr Gly Arg Glu Arg Asp
35 40 45

Gly Pro Pro Pro Arg Arg Val Phe Gly Gly Gly Leu Arg Val Gly Asp
50 55 60

Pro Ser Gly Tyr Asp Arg Gly Glu Ser Lys Pro Ile Gly Gly Pro Leu
65 70 75 80

Thr Glu Lys Arg Ser Asp Trp His Ser Tyr Phe Met Arg Ile Ala Gly
85 90 95

Glu Val Ala Thr Arg Ala Thr Cys Asp Arg Lys His Val Gly Ala Val
100 105 110

Ile Val Arg Asn Arg Thr Ile Leu Ser Thr Gly Tyr Asn Gly Ser Ile
115 120 125

Arg Gly Met Pro His Cys Asp Asp Val Gly His Asp Met Val Asp Gly
130 135 140

```
His Cys Ile Ala Thr Ile His Ala Glu Ala Asn Ala Ile Leu Gln Ala
145                 150                 155                 160


Ala Arg Asn Gly Val Met Ile Gln Asp Gly Ser Ile Tyr Ile Thr Ala
                165                 170                 175


Ser Pro Cys Trp Asn Cys Phe Lys Leu Val Ala Asn Ala Gly Leu Lys
            180                 185                 190


Arg Val Tyr Tyr Gly Glu Phe Tyr Arg Asp Lys Arg Ser Phe Glu Val
        195                 200                 205


Ala Arg Arg Leu Gly Ile Asp Leu Met His Ile Glu Val
    210                 215                 220


<210>  179
<211>  155
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.70;
       tr|A0A1B8WPS3|A0A1B8WPS3_9BACI polypeptide sequence

<400>  179

Met Glu Gly Val Gln Leu Ile Tyr Gln Phe Gln Trp Gly Asn Leu Ile
1               5                   10                  15


Met Thr Val Asn Lys Glu Asp Leu Tyr Leu Ile Asp Val Ala Arg Asn
            20                  25                  30


Thr Ile Lys Thr Leu Tyr Val Asp Gly Lys His His Val Gly Ala Ala
        35                  40                  45


Val Arg Thr Lys Thr Gly Lys Ile Tyr Ser Ala Val His Leu Glu Ala
    50                  55                  60


Asn Ile Gly Arg Val Ser Val Cys Ala Glu Ala Ile Ala Leu Gly Lys
65                  70                  75                  80


Ala Ile Ser Glu Gly Glu Ser Glu Phe Asp Thr Ile Val Ala Val Arg
                85                  90                  95
```

```
His Pro Asp Pro Thr Gln Glu Asn Gln Lys Ile Glu Val Val Ser Pro
            100                 105                 110


Cys Gly Ile Cys Arg Glu Leu Ile Ser Asp Tyr Gly Lys Gly Thr Asn
        115                 120                 125


Val Ile Leu Lys Asn Lys Glu Gly Tyr Ile Lys Thr Val Ile Ser Asp
    130                 135                 140


Leu Leu Pro Asn Lys Tyr Ile Arg Glu Asp Asn
145                 150                 155


<210>  180
<211>  156
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: pYY-BEM4.71;
       tr|A0A1W5ZQK9|A0A1W5ZQK9_9BACI polypeptide sequence

<400>  180

Met Asn Arg Phe Met Glu Arg Ala Val Ser Leu Ala Ala Glu Asn Val
1               5                   10                  15


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


Leu Val Ala Glu Gly Val Asn Glu Met His Leu Asn Tyr Asp Val Ser
        35                  40                  45


Gly His Ala Glu Leu Leu Ala Ile Arg Arg Ala Gln Gly Glu Leu Gln
    50                  55                  60


Thr His Asp Leu Ser Gly Tyr Thr Met Tyr Ala Ser Gly Glu Pro Cys
65                  70                  75                  80


Pro Met Cys Leu Ser Ala Met Tyr Phe Ala Gly Ile Lys Asp Val Phe
                85                  90                  95


Tyr Cys Ala Thr Val Glu Glu Ala Ala Gln Val Gly Leu Glu Lys Ser
```

```
            100                 105                 110

Lys Asn Val Tyr Asp Asp Leu Gln Lys Ser Lys Gly Glu Arg Ser Leu
        115                 120                 125

Val Met Lys Gln Met Pro Leu Glu Asp Asp Gln Glu Asp Pro Met Lys
    130                 135                 140

Leu Trp Asp Glu Arg Thr Asn His Asn Gly Thr Ser
145                 150                 155
```

<210> 181
<211> 340
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.72;
tr|A0A378V0W4|A0A378V0W4_MYCFO polypeptide sequence

<400> 181

```
Met Val His Ala Gln Phe Asp Pro Thr Ala Arg Gln Ala Leu Ala Ala
1               5                   10                  15

Thr Ala Val Glu Ala Lys Thr Arg Lys Asp Leu Thr Trp Gln Gln Ile
            20                  25                  30

Ala Asp Ala Ala Glu Leu Ser Pro Ala Phe Val Thr Ala Ala Val Leu
        35                  40                  45

Gly Gln His Ala Leu Pro Ala Arg Ser Ala Glu Ala Val Ala Ala Leu
    50                  55                  60

Leu Gly Leu Asp Asp Asp Ala Ala Leu Leu Leu Gln Thr Ile Pro Ile
65                  70                  75                  80

Arg Gly Ser Ile Pro Gly Gly Ile Pro Thr Asp Pro Thr Ile Tyr Arg
                85                  90                  95

Phe Tyr Glu Met Leu Gln Val Tyr Gly Thr Thr Leu Lys Ala Leu Val
            100                 105                 110
```

```
His Glu Gln Phe Gly Asp Gly Ile Ile Ser Ala Ile Asn Phe Lys Leu
        115                 120                 125

Asp Val Arg Lys Val Ala Asp Pro Glu Gly Gly Glu Arg Ala Val Ile
    130                 135                 140

Thr Leu Asp Gly Lys Tyr Leu Pro Pro Asn Pro Phe Asp Arg Val Arg
145                 150                 155                 160

Tyr Arg Gly Gly Leu Met Asp Phe Ala Gln Arg Thr Ile Asp Ile Ala
                165                 170                 175

Arg Gln Asn Val Ala Glu Gly Gly Arg Pro Phe Ala Thr Val Ile Val
            180                 185                 190

Lys Asn Gly Glu Ile Leu Ala Glu Ser Pro Asn Leu Val Ala Gln Thr
        195                 200                 205

His Asp Pro Thr Ala His Ala Glu Ile Leu Ala Ile Arg Lys Ala Cys
    210                 215                 220

Thr Arg Ile Gly Thr Glu His Leu Ile Gly Ala Thr Ile Tyr Val Leu
225                 230                 235                 240

Ala Gln Pro Cys Pro Met Cys Leu Gly Ser Leu Tyr Tyr Cys Ser Pro
                245                 250                 255

Asp Glu Val Val Phe Leu Thr Thr Arg Asp Ala Tyr Glu Pro His Tyr
            260                 265                 270

Val Asp Asp Arg Lys Tyr Phe Glu Leu Asn Met Phe Tyr Asp Glu Phe
        275                 280                 285

Ala Lys Pro Trp Asp Gln Arg Arg Leu Pro Met Arg Tyr Glu Pro Arg
    290                 295                 300

Asp Ala Ala Val Asp Val Tyr Lys Leu Trp Gln Glu Arg Asn Gly Gly
305                 310                 315                 320

Glu Arg Arg Val Pro Gly Ala Pro Thr Ser Thr Arg Pro Gly Lys Asn
                325                 330                 335
```

Pro Arg Gly Glu
340

<210> 182
<211> 157
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.73;
tr|I3XF03|I3XF03_RHIFR polypeptide sequence

<400> 182

Met Lys Gln Arg Cys Met Ser Pro Lys Ser Ala Gln Arg Phe Trp Asp
1 5 10 15

Asn Asp Met His Asn Asn Lys Asp Arg Pro Met Ser Glu Asn Glu Leu
20 25 30

Phe Val Ala Ala Arg Glu Ala Met Ala Lys Ala His Ala Pro Tyr Ser
35 40 45

Lys Phe Pro Val Gly Ala Ala Ile Arg Ala Glu Asp Gly Gln Ile Tyr
50 55 60

Thr Gly Ala Asn Ile Glu Asn Leu Ser Phe Pro Glu Gly Trp Cys Ala
65 70 75 80

Glu Thr Thr Ala Ile Ser His Met Val Met Ala Gly Gln Arg Lys Ile
85 90 95

Met Glu Val Ala Val Ile Ala Glu Lys Leu Ala Leu Cys Pro Pro Cys
100 105 110

Gly Gly Cys Arg Gln Arg Leu Ala Glu Phe Ser Gly Ala Ser Thr Arg
115 120 125

Ile Tyr Leu Cys Asp Glu Thr Gly Ile Lys Lys Ser Leu Ala Leu Ser
130 135 140

Asp Leu Leu Pro His Ser Phe Glu Thr Glu Ile Leu Gly

145 150 155

<210> 183
<211> 595
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.74;
tr|F8IEF3|F8IEF3_ALIAT polypeptide sequence

<400> 183

Met Asp Ala Lys Glu Leu Glu Thr Arg Gly Trp Leu Cys Met Arg Ala
1 5 10 15

Val Asp Val Ile Asp Lys Lys Arg Arg Gly Glu Ala Leu Ala Glu Glu
20 25 30

Glu Leu Arg Phe Leu Ile Glu Gly Tyr Val Ala Gly Arg Ile Pro Asp
35 40 45

Tyr Gln Met Ser Ala Phe Leu Met Ala Val Val Trp Arg Gly Met Thr
50 55 60

Arg Glu Glu Thr Leu Val Leu Thr Arg Leu Leu Ala Asp Ser Gly Glu
65 70 75 80

Arg Leu Asp Leu Ser Gly Ile Pro Gly Val Lys Val Asp Lys His Ser
85 90 95

Thr Gly Gly Val Gly Asp Lys Ala Thr Leu Val Val Leu Pro Leu Val
100 105 110

Ala Ser Ile Gly Val Pro Val Ile Lys Met Ser Gly Arg Gly Leu Gly
115 120 125

His Thr Gly Gly Thr Ile Asp Lys Leu Glu Ser Ile Pro Gly Phe Arg
130 135 140

Thr Asp Leu Ser Val Ala Glu Leu Val Ala Gln Val Arg Gln Val Gly
145 150 155 160

```
Ile Ala Leu Gly Gly Gln Thr Ala Asp Leu Ala Pro Ala Asp Lys Lys
                165                 170                 175

Leu Tyr Ala Leu Arg Asp Val Thr Gly Thr Val Glu Ser Leu Pro Leu
            180                 185                 190

Ile Ala Ser Ser Val Met Ser Lys Lys Leu Ala Gly Gly Ala Asp Ala
        195                 200                 205

Ile Val Leu Asp Val Lys Val Gly Asp Gly Ala Phe Met Lys Ser Arg
    210                 215                 220

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

Ala Gly Arg Arg Thr Val Ala Val Leu Ser Asn Met Asp Gln Pro Leu
                245                 250                 255

Gly Cys Ala Ile Gly Asn Ala Leu Glu Val Ala Glu Ala Ile Arg Val
            260                 265                 270

Leu Ser Gly Glu Gly Pro Phe Asp Leu Ala Glu Ile Ala Leu Ala Leu
        275                 280                 285

Ala Glu Glu Met Thr Val Leu Ala Gly Val Ala Ala Thr Arg Glu Glu
    290                 295                 300

Ala Arg Arg Met Leu Arg Gln Ser Val Ala Glu Gly Arg Ala Leu Glu
305                 310                 315                 320

Thr Leu Arg Arg Trp Ile Ala Ala Gln Gly Gly Asp Pro Ala Val Val
                325                 330                 335

Asp Asp Pro Ser Arg Leu Pro Gln Ala Pro Val Gln Met Pro Tyr Leu
            340                 345                 350

Pro Lys Lys Ala Gly Phe Val Ala Lys Leu Ser Ala Leu Ala Phe Gly
        355                 360                 365

Leu Ala Ala Met Arg Leu Gly Ala Gly Arg Glu Thr Lys Glu Glu Ala
    370                 375                 380
```

```
Ile Asp Pro Ser Val Gly Ile Val Leu His Ala Lys Val Gly Asp Arg
385                 390                 395                 400

Val Gln Thr His Arg Pro Met Phe Thr Val His Ala Arg Thr Gly Glu
                405                 410                 415

Asp Ala Leu Arg Cys Ile Gln Glu Leu Glu Ala Ala Ile Gln Ile Ser
            420                 425                 430

Asp Asp Pro Val Glu Ala Pro Pro Leu Ile Leu Ala Arg Ile Asp Arg
        435                 440                 445

Ser Glu Ala Leu Pro Tyr Ala Asp Leu Met Asp Ala Ala Arg Glu Ala
    450                 455                 460

Arg Asp Arg Ala Tyr Val Pro Tyr Ser Gly Phe Ala Val Gly Ala Ala
465                 470                 475                 480

Leu Glu Leu Ala Asp Gly Arg Met Val Thr Gly Ala Asn Val Glu Asn
                485                 490                 495

Ala Ser Tyr Gly Leu Thr Asn Cys Ala Glu Arg Ser Ala Val Phe Arg
            500                 505                 510

Ala Val Ala Glu Gly Gly Pro Gly Thr Lys Pro Glu Ile Arg Ala Val
        515                 520                 525

Ala Val Ile Ala Asp Ser Pro Glu Pro Val Ser Pro Cys Gly Ala Cys
    530                 535                 540

Arg Gln Val Leu Ala Glu Phe Cys Ser Pro Asp Thr Pro Val Tyr Leu
545                 550                 555                 560

Gly Asn Leu Gln Gly Asp Val Arg Glu Thr Thr Val Gly Ala Leu Leu
                565                 570                 575

Pro Gly Ala Phe Thr Asp Ala Gln Met Ala Asn Val Arg Arg Gln Asp
            580                 585                 590
```

Lys Glu Ala
595

<210> 184
<211> 153
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.75;
tr|A0A1G3M638|A0A1G3M638_9SPIR polypeptide sequence

<400> 184

Met Lys Thr Thr Asn Ile Asn Ala Leu Asp Lys Trp Asp Leu Arg Phe
1 5 10 15

Leu Gln Met Ala Glu His Val Ala Glu Trp Ser Lys Asp Pro Ser Thr
20 25 30

Lys Val Gly Ala Val Ile Val Arg Pro Asp Arg Thr Ile Ala Ser Val
35 40 45

Gly Phe Asn Gly Phe Ala Arg Gly Val Arg Asp Thr Val Glu Arg Leu
50 55 60

Trp Asn Arg Glu Leu Lys Tyr Pro Leu Thr Val His Ala Glu Leu Asn
65 70 75 80

Ala Ile Leu Ser Ala His Glu Pro Val Arg Gly His Ser Leu Tyr Val
85 90 95

Ser Pro Leu Ser Pro Cys Ser Asn Cys Ala Gly Val Ile Ile Gln Ser
100 105 110

Gly Ile Ala Arg Val Val Ala Lys Cys Gly Gln Val Asn Asn Pro Ala
115 120 125

Gln Trp Ser Glu Ser Phe Asn Leu Ala Leu Thr Ala Phe Ala Glu Ala
130 135 140

Gly Val Ser Val Ile Leu Val Glu His
145 150

<210> 185
<211> 177
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.76;
tr|A0A3D9LFR2|A0A3D9LFR2_9MICC polypeptide sequence

<400> 185

Met Glu Gln Asn Asp His Gly Ser Ser Gly Ala Phe Ser Asp Pro Phe
1 5 10 15

Glu Asp Asp Ile Pro Leu Thr Ala Ser Leu Pro Arg Ile Thr Gly Thr
20 25 30

Gly Ser Gly Ile Asp Trp Gln Arg Leu Glu Ser Thr Ala Arg Ala Ala
35 40 45

Met Thr Arg Ala Tyr Val Pro Tyr Ser Arg Phe Pro Val Gly Ala Ala
50 55 60

Ala Leu Val Glu Asp Gly Arg Val Val Ala Gly Cys Asn Ile Glu Asn
65 70 75 80

Ala Ser Leu Gly Leu Thr Leu Cys Ala Glu Cys Ser Leu Val Ser Asn
85 90 95

Leu Gln Met Ser Gly Gly Gly Arg Ile Val Ala Phe Tyr Cys Val Asp
100 105 110

Gly Asn Gly Glu Val Leu Met Pro Cys Gly Arg Cys Arg Gln Leu Leu
115 120 125

Tyr Glu Phe His Ala Pro Gly Met Arg Leu Met Gly Pro Asp Gly Glu
130 135 140

Leu Thr Met Asp Glu Val Leu Pro Leu Ala Phe Gly Pro Ala Asp Met
145 150 155 160

Thr His Leu Ser Asp Ser Ala Ala Ser Thr Asp Asp Pro Gly Arg Thr
165 170 175

Arg

<210> 186
<211> 172
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.77;
tr|A0A3B9YGB5|A0A3B9YGB5_9BACT polypeptide sequence

<400> 186

Met Ala Lys Pro Ile Ser Lys Lys Tyr Arg Lys Leu Ile Glu Thr Ala
1 5 10 15

Lys Ala Ala Arg Lys Lys Ala Tyr Ser Pro Tyr Ser Arg Tyr Gln Val
20 25 30

Gly Ala Ala Val Leu Thr Glu Ser Gly Arg Ile Tyr Ser Gly Ala Asn
35 40 45

Met Glu Asn Ala Ser Tyr Gly Leu Cys Met Cys Ala Glu Arg Val Ala
50 55 60

Ile Ala Asn Ala Val Thr Arg Gly Glu Lys Val Leu Gln Ala Val Cys
65 70 75 80

Val Val Gly Lys Lys Ala Arg Pro Cys Gly Ala Cys Arg Gln Val Met
85 90 95

Leu Glu Phe Ser Thr Lys Glu Thr Glu Leu Leu Met Val Asp Ile Asp
100 105 110

Pro Asn Ala Arg Arg Asp Thr Val Ile Arg Thr Arg Val Tyr Ser Met
115 120 125

Leu Pro Asn Pro Phe Asp Pro Phe Glu Ser Gly Met Leu Pro Gln His
130 135 140

Pro Gln Asn Leu Leu Arg Arg Arg Lys Ser Pro Gln Pro Arg Arg Lys

145 150 155 160

Arg Arg Ser Arg Pro Val His Arg Glu Val Ser Arg
165 170

<210> 187
<211> 391
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.78;
tr|A0A182F569|A0A182F569_ANOAL polypeptide sequence

<400> 187

Met Pro Arg Pro Ser Gln Phe Arg Val Ser Ser Ser Gln Ser Leu Ser
1 5 10 15

Asn Ser Gln Ile Gln Ala Ser Gln Ser Ser Asp Ser Val Val Asp Ile
20 25 30

Thr Ser Tyr Val Asn Ala Val Val Lys Ala Leu Leu Asn Leu Ser Cys
35 40 45

Thr Lys Thr Ile Ile Lys Arg Ala Asp Leu Val Asn Ile Ala Leu Lys
50 55 60

Gly Asn Gly Arg Leu Ile Gly Arg Val Leu Gln Asp Ala Asn Ile Glu
65 70 75 80

Leu Lys Glu Ile Tyr Gly Tyr Glu Leu Ile Glu Val Glu Lys Ser Lys
85 90 95

Thr Met Ile Leu Cys Ser Thr Leu Ala Ala Gly Ser Met Asp Glu Leu
100 105 110

Asn Asp Ala Asn Arg Arg Arg Tyr Thr Phe Leu Tyr Leu Ile Leu Gly
115 120 125

Tyr Ile Phe Met Lys Asn Gly Ser Val Pro Glu Thr Ile Val Trp Glu
130 135 140

```
Phe Leu Glu Thr Leu Gly Ile Glu Glu Gln Gln Glu His Asn Tyr Phe
145                 150                 155                 160

Gly Asp Val Arg Lys Leu Tyr Asp Ser Leu Phe Lys Gln Ala Tyr Leu
                165                 170                 175

Thr Arg Thr Lys Gln Ala Leu Glu Gly Leu Asn Asp Asp Val Met Leu
            180                 185                 190

Ile Ser Trp Gly Val Arg Ser Lys His Glu Val Ser Lys Lys Asp Ile
        195                 200                 205

Leu Ala Gly Phe Cys Lys Val Met Asn Arg Asp Pro Val Asp Phe Lys
    210                 215                 220

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

Ile Asn Gly Thr Val Asp Gly Arg Asn Thr Val Glu Tyr Ser Ser Leu
                245                 250                 255

Asp Ala Ser Val Lys Glu Leu Ile Glu Ala Ala Ile Lys Val Arg Asn
            260                 265                 270

Asn Ala Tyr Cys Pro Tyr Ser Asn Phe Ala Val Gly Ala Ala Leu Arg
        275                 280                 285

Thr Val Gly Gly Asp Ile Val Thr Gly Cys Asn Val Glu Asn Gly Thr
    290                 295                 300

Phe Gly Pro Ser Val Cys Ala Glu Arg Thr Ala Val Cys Lys Ala Val
305                 310                 315                 320

Ser Glu Gly His Arg Glu Phe Thr Ala Val Ala Val Val Ala Phe Gln
                325                 330                 335

Glu Thr Glu Phe Thr Ala Pro Cys Gly Thr Cys Arg Gln Thr Leu Ser
            340                 345                 350

Glu Phe Ser Arg Lys Asp Ile Pro Ile Tyr Leu Val Lys Pro Ser Pro
        355                 360                 365
```

Val Arg Val Met Val Thr Ser Leu Phe Gln Leu Leu Pro His Ala Phe
    370   375   380

Ser Pro Ser Phe Leu Asn Lys
385   390

<210> 188
<211> 151
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.79;
       tr|A0A264Z0D4|A0A264Z0D4_9BACI polypeptide sequence

<400> 188

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

Ala Tyr Val Gln Tyr Ser Asn Phe His Val Gly Ala Ala Leu Leu Thr
            20   25   30

Lys Asp Gly Lys Leu Tyr His Gly Cys Asn Ile Glu Asn Ala Ser Tyr
        35   40   45

Gly Leu Thr Asn Cys Ala Glu Arg Thr Ala Ile Phe Lys Ala Val Ser
    50   55   60

Glu Gly Glu Lys Glu Phe Gln Ala Ile Ala Val Val Gly Asp Thr Glu
65   70   75   80

Gly Pro Ile Ser Pro Cys Gly Ala Cys Arg Gln Val Leu Ala Glu Phe
                85   90   95

Phe Ser Pro Asp Thr Val Val Ile Leu Ala Asn Leu Lys Gly Asp His
            100   105   110

Val Val Thr Asn Ile Asn Glu Leu Leu Pro Gly Phe Phe Ser Ser Lys
        115   120   125

Asp Leu Gln Lys Lys Val Lys Asn Cys Phe Glu Lys Asn Ala Leu Gly

130 135 140

Ser Ser Cys Leu Arg Pro Ile
145 150

<210> 189
<211> 229
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pYY-BEM4.80;
tr|A0A1L9Q1R3|A0A1L9Q1R3_ASPVE polypeptide sequence

<400> 189

Met Pro Leu Ser Ala Glu Glu Ala Ala Leu Val Glu Thr Ala Thr Ala
1 5 10 15

Thr Ile Asn Ser Ile Pro Leu Ser Glu Asp Tyr Ser Val Ala Ser Ala
20 25 30

Ala Lys Ala Ser Asp Gly Arg Val Phe Thr Gly Val Asn Val Tyr His
35 40 45

Phe Thr Gly Gly Pro Cys Ala Glu Leu Val Val Leu Gly Val Ala Ala
50 55 60

Ala Ala Gly Ala Ala Gln Leu Thr His Ile Val Ala Val Ala Asn Glu
65 70 75 80

Gln Arg Gly Ile Leu Ser Pro Cys Gly Arg Cys Arg Gln Val Leu Leu
85 90 95

Asp Leu Gln Pro Asn Ile Gln Val Ile Val Gly Lys Glu Gly Ser Glu
100 105 110

Gln Ser Val Pro Val Ala Gln Leu Leu Pro Phe Ser Tyr Arg Gln Pro
115 120 125

Asp Gln His Thr Pro Val Ile Phe Lys Ala Leu Thr Ser Ser Gly Pro
130 135 140

```
Val Val Val Asp Phe Phe Ala Thr Trp Cys Gly Pro Cys Lys Ala Val
145                 150                 155                 160


Ala Pro Val Val Gly Lys Leu Ser Glu Thr Tyr Thr Asp Val Arg Phe
                165                 170                 175


Ile Gln Val Asp Val Asp Lys Ala Arg Ser Ile Ser Gln Glu His Asp
            180                 185                 190


Ile Arg Ala Met Pro Thr Phe Val Leu Tyr Lys Asp Gly Lys Leu Leu
        195                 200                 205


Asp Lys Arg Val Val Gly Gly Asn Met Lys Glu Leu Glu Glu Gln Ile
    210                 215                 220


Lys Ala Ile Ile Ala
225
```

<210> 190
<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide

<400> 190

```
Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg
1               5                   10                  15


Lys Val
```

<210> 191
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide

<400> 191

```
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1               5                   10                  15
```

<210> 192
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide

<400> 192

Lys Lys Thr Glu Leu Gln Thr Thr Asn Ala Glu Asn Lys Thr Lys Lys
1 5 10 15

Leu

<210> 193
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide

<400> 193

Lys Arg Gly Ile Asn Asp Arg Asn Phe Trp Arg Gly Glu Asn Gly Arg
1 5 10 15

Lys Thr Arg

<210> 194
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide

<400> 194

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys
1 5 10 15

<210> 195
<211> 7

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide

<400> 195

Pro Lys Lys Lys Arg Lys Val
1 5

<210> 196
<211> 30
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide

<400> 196

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
20 25 30

<210> 197
<211> 1368
<212> PRT
<213> Streptococcus pyogenes

<220>
<221> MISC_FEATURE
<222> (1)..(1368)
<223> Exemplary PAM-binding SpCas9 polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (5)..(58)
<223> RuvC domain

<220>
<221> MISC_FEATURE
<222> (718)..(764)
<223> RuvC domain

<220>
<221> MISC_FEATURE
<222> (773)..(902)

<223> HNH domain

<400> 197

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                  10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60


Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80


Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95


Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110


His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125


His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140


Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160


Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175


Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190


Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
```

```
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
```

```
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
```

```
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
               645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
               725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
               805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
```

```
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065
```

```
Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260
```

```
His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365
```

<210> 198
<211> 4104
<212> DNA
<213> Streptococcus pyogenes

<220>
<221> misc_feature
<222> (1)..(4104)
<223> wild type spCas9 polynucleotide sequence

<400> 198

```
atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg       60

atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc      120

cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa      180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt      240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga      300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga      360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa      420
```

aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct 600

attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat 720

ctcattgctt tgtcattggg attgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatagt gaaataacta aggctcccct atcagcttca 960

atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agataggggg atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta 2160

catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggtat tttacagact 2220

gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt 2280

gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg 2340

aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt 2400

gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac 2460

atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt 2520

gttccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat 2580

aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac 2640

tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg 2700

aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg 2760

gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact 2820

aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa 2880

ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac 2940

catcatgccc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat 3000

ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg 3060

attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat 3120

atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct 3180

ctaatcgaaa ctaatgggga aactggagaa attgtctggg ataaagggcg agattttgcc 3240

acagtgcgca aagtattgtc catgccccaa gtcaatattg tcaagaaaac agaagtacag 3300

acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360

cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420

tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480

gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540

ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600

agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660

aaaggaaatg agctggctct gccaagcaaa tatgtgaatt ttttatattt agctagtcat 3720

tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780

cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840

ttagcagatg ccaatttaga taaagttctt agtgcatata acaaacatag agacaaacca 3900

atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960

gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020

gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080

ttgagtcagc taggaggtga ctga 4104

<210> 199
<211> 1367
<212> PRT
<213> Streptococcus pyogenes

<220>
<221> MISC_FEATURE
<222> (1)..(1367)
<223> spCas9 polypeptide sequence

<400> 199

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

```
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
```

```
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
```

```
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705                 710                 715                 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
```

```
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
        755                 760                 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
    770                 775                 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785                 790                 795                 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
                805                 810                 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
            820                 825                 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
        835                 840                 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
    850                 855                 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865                 870                 875                 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
                885                 890                 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
            900                 905                 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
        915                 920                 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
    930                 935                 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945                 950                 955                 960
```

```
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
               965                 970                 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
           980                 985                 990

Gly Thr Ala Leu Ile Lys Lys Tyr  Pro Lys Leu Glu Ser  Glu Phe Val
        995                1000                1005

Tyr Gly  Asp Tyr Lys Val Tyr  Asp Val Arg Lys Met  Ile Ala Lys
    1010                1015                1020

Ser Glu  Gln Glu Ile Gly Lys  Ala Thr Ala Lys Tyr  Phe Phe Tyr
    1025                1030                1035

Ser Asn  Ile Met Asn Phe Phe  Lys Thr Glu Ile Thr  Leu Ala Asn
    1040                1045                1050

Gly Glu  Ile Arg Lys Arg Pro  Leu Ile Glu Thr Asn  Gly Glu Thr
    1055                1060                1065

Gly Glu  Ile Val Trp Asp Lys  Gly Arg Asp Phe Ala  Thr Val Arg
    1070                1075                1080

Lys Val  Leu Ser Met Pro Gln  Val Asn Ile Val Lys  Lys Thr Glu
    1085                1090                1095

Val Gln  Thr Gly Gly Phe Ser  Lys Glu Ser Ile Leu  Pro Lys Arg
    1100                1105                1110

Asn Ser  Asp Lys Leu Ile Ala  Arg Lys Lys Asp Trp  Asp Pro Lys
    1115                1120                1125

Lys Tyr  Gly Gly Phe Asp Ser  Pro Thr Val Ala Tyr  Ser Val Leu
    1130                1135                1140

Val Val  Ala Lys Val Glu Lys  Gly Lys Ser Lys Lys  Leu Lys Ser
    1145                1150                1155

Val Lys  Glu Leu Leu Gly Ile  Thr Ile Met Glu Arg  Ser Ser Phe
    1160                1165                1170
```

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

Val Lys  Lys Asp Leu Ile Ile  Lys Leu Pro Lys Tyr  Ser Leu Phe
    1190                1195                1200

Glu Leu  Glu Asn Gly Arg Lys  Arg Met Leu Ala Ser  Ala Gly Glu
    1205                1210                1215

Leu Gln  Lys Gly Asn Glu Leu  Ala Leu Pro Ser Lys  Tyr Val Asn
    1220                1225                1230

Phe Leu  Tyr Leu Ala Ser His  Tyr Glu Lys Leu Lys  Gly Ser Pro
    1235                1240                1245

Glu Asp  Asn Glu Gln Lys Gln  Leu Phe Val Glu Gln  His Lys His
    1250                1255                1260

Tyr Leu  Asp Glu Ile Ile Glu  Gln Ile Ser Glu Phe  Ser Lys Arg
    1265                1270                1275

Val Ile  Leu Ala Asp Ala Asn  Leu Asp Lys Val Leu  Ser Ala Tyr
    1280                1285                1290

Asn Lys  His Arg Asp Lys Pro  Ile Arg Glu Gln Ala  Glu Asn Ile
    1295                1300                1305

Ile His  Leu Phe Thr Leu Thr  Asn Leu Gly Ala Pro  Ala Ala Phe
    1310                1315                1320

Lys Tyr  Phe Asp Thr Thr Ile  Asp Arg Lys Arg Tyr  Thr Ser Thr
    1325                1330                1335

Lys Glu  Val Leu Asp Ala Thr  Leu Ile His Gln Ser  Ile Thr Gly
    1340                1345                1350

Leu Tyr  Glu Thr Arg Ile Asp  Leu Ser Gln Leu Gly  Gly Asp
    1355                1360                1365
```

<210> 200
<211> 4212
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: wild-type Cas9 polynucleotide sequence

<220>
<221> misc_feature
<222> (5)..(58)
<223> RuvC domain

<220>
<221> misc_feature
<222> (718)..(764)
<223> RuvC domain

<220>
<221> misc_feature
<222> (772)..(901)
<223> HNH domain

<220>
<221> misc_feature
<222> (906)..(1101)
<223> RuvC domain

<400> 200
atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60

ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120

cattcgatta aaaagaatct tatcggtgcc ctcctattcg atagtggcga aacggcagag 180

gcgactcgcc tgaaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240

tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300

ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360

aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420

aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480

atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540

gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600

ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660

cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac 720

cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780

gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840

caaattggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc 900

ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca 960

atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt 1020

cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca 1080

ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta 1140

gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcgaaga tctactgcga 1200

aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat 1260

gctatactta gaaggcagga ggatttttat ccgttcctca aagacaatcg tgaaaagatt 1320

gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct 1380

cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa 1440

gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag 1500

aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagtg 1560

tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta 1620

agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca 1680

gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc 1740

tccggggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata 1800

attaaagata aggacttcct ggataacgaa gagaatgaag atatcttaga agatatagtg 1860

ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct 1920

cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgctatac gggctgggga 1980

cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc 2040

gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac 2100

tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg 2160

cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca 2220

gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaaccgga aaacattgta 2280

atcgagatgg cacgcgaaaa tcaaacgact cagaaggggc aaaaaaacag tcgagagcgg 2340

atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaaa ggagcatcct 2400

gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg 2460

gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac 2520

attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgct tacacgctcg 2580

gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag 2640

aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagtt cgataactta 2700

actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggatttat taaacgtcag 2760

ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820

acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880

aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940

taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000

tacccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060

atgatcgcga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120

aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180

cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240

gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg 3300

cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc 3360

gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420

tattctgtcc tagtagtggc aaaagttgag aagggaaaat ccaagaaact gaagtcagtc 3480

aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540

ttccttgagg cgaaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600

tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660

caaaagggga acgaactcgc actaccgtct aaatacgtga atttcctgta tttagcgtcc 3720

cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact ttttgttgag 3780

cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840

atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900

cccatacgtg agcaggcgga aaatattatc catttgttta ctcttaccaa cctcggcgct 3960

ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020

gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080

gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140

tacaaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200

aaggctgcag ga 4212

<210> 201
<211> 1368
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: exemplary catalytically active Cas9 nickase (nCas9) polypeptide sequence

<400> 201

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

```
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
```

```
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560
```

```
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
```

```
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990
```

```
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080


Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095


Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110


Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125


Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140


Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155


Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170


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


Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
```

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 202
<211> 4107
<212> DNA
<213> Streptococcus pyogenes

<220>

<221> misc_feature
<222> (1)..(4107)
<223> Cas9 from Streptococcus pyogenes (NCBI Ref. Seq.: NC_002737.2)
polynucleotide sequence

<400> 202
atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600

attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720

ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960

atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta 2160

catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220

gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt 2280

attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt 2340

atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct 2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460

gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac 2520

attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580

gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640

aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700

acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa 2760

ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820

actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct 2880

aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940

taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000

tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060

atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120

aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180

cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240

gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300

cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360

gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420

tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480

aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540

tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600

tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660

caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttttata tttagctagt 3720

cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780

cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840

attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900

ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960

cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020

gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080

gatttgagtc agctaggagg tgactga 4107

<210> 203
<211> 1368
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of sequence: catalytically inactive Cas9 (dCas9) polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (5)..(58)
<223> RuvC domain

<220>
<221> MISC_FEATURE

<222> (718)..(764)
<223> RuvC domain

<220>
<221> MISC_FEATURE
<222> (773)..(902)
<223> HNH domain

<220>
<221> MISC_FEATURE
<222> (907)..(1102)
<223> RuvC domain

<400> 203

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145   150   155   160

```
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
```

```
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
```

```
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
```

```
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020
```

```
Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
```

```
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350

Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365


<210>  204
<211>  345
<212>  PRT
<213>  Sulfolobus islandicus


<220>
<221>  MISC_FEATURE
<222>  (1)..(345)
<223>  tr|F0NN87|F0NN87_SULIHCRISPR-associatedCasx protein OS =
       Sulfolobus islandicus (strain HVE10/4) GN = SiH_0402 PE=4 SV=1);
       CasX polypeptide sequence

<400>  204
```

```
Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1               5                  10                  15

Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
            20                  25                  30

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

Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
    50                  55                  60

Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65                  70                  75                  80

Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
                85                  90                  95

Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
            100                 105                 110

Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
        115                 120                 125

Pro Glu Phe Tyr Glu Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
    130                 135                 140

Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Ile Ala Ala
145                 150                 155                 160

Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
                165                 170                 175

Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
            180                 185                 190

Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
        195                 200                 205

Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
    210                 215                 220
```

```
Asn Pro Asn Val Ser Val Val Arg Ile Tyr Thr Ile Ser Asp Ala Val
225                 230                 235                 240


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


Lys Leu Leu Glu Lys Arg Tyr Leu Leu Ser Glu Arg Leu Glu Ala Ile
            260                 265                 270


Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
        275                 280                 285


Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
    290                 295                 300


Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305                 310                 315                 320


Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
                325                 330                 335


Asn Gly Glu Leu Ile Arg Gly Glu Gly
            340                 345


<210>  205
<211>  345
<212>  PRT
<213>  Sulfolobus islandicus


<220>
<221>  MISC_FEATURE
<222>  (1)..(345)
<223>  tr|F0NH53|F0NH53_SULIR CRISPR associated protein, Casx OS =
       Sulfolobus islandicus (strain REY15A) GN=SiRe_0771 PE=4 SV=1);
       CasX polypeptide sequence

<400>  205

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1               5                   10                  15


Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
```

```
            20                  25                  30

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

Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
    50                  55                  60

Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65                  70                  75                  80

Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
                85                  90                  95

Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
            100                 105                 110

Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
        115                 120                 125

Pro Glu Phe Tyr Lys Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
    130                 135                 140

Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Met Ala Ala
145                 150                 155                 160

Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
                165                 170                 175

Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
            180                 185                 190

Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
        195                 200                 205

Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
    210                 215                 220

Asn Pro Asn Val Ser Val Val Ser Ile Tyr Thr Ile Ser Asp Ala Val
225                 230                 235                 240
```

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

Lys Leu Leu Glu Lys Arg Asp Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270

Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285

Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300

Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320

Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335

Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 206
<211> 987
<212> PRT
<213> Deltaproteobacteria

<220>
<221> MISC_FEATURE
<222> (1)..(987)
<223> CasX polypeptide sequence

<400> 206

Met Glu Lys Arg Ile Asn Lys Ile Arg Lys Lys Leu Ser Ala Asp Asn
1 5 10 15

Ala Thr Lys Pro Val Ser Arg Ser Gly Pro Met Lys Thr Leu Leu Val
20 25 30

Arg Val Met Thr Asp Asp Leu Lys Lys Arg Leu Glu Lys Arg Arg Lys
35 40 45

```
Lys Pro Glu Val Met Pro Gln Val Ile Ser Asn Asn Ala Ala Asn Asn
    50                  55                  60

Leu Arg Met Leu Leu Asp Asp Tyr Thr Lys Met Lys Glu Ala Ile Leu
65                  70                  75                  80

Gln Val Tyr Trp Gln Glu Phe Lys Asp Asp His Val Gly Leu Met Cys
                85                  90                  95

Lys Phe Ala Gln Pro Ala Ser Lys Lys Ile Asp Gln Asn Lys Leu Lys
            100                 105                 110

Pro Glu Met Asp Glu Lys Gly Asn Leu Thr Thr Ala Gly Phe Ala Cys
        115                 120                 125

Ser Gln Cys Gly Gln Pro Leu Phe Val Tyr Lys Leu Glu Gln Val Ser
    130                 135                 140

Glu Lys Gly Lys Ala Tyr Thr Asn Tyr Phe Gly Arg Cys Asn Val Ala
145                 150                 155                 160

Glu His Glu Lys Leu Ile Leu Leu Ala Gln Leu Lys Pro Val Lys Asp
                165                 170                 175

Ser Asp Glu Ala Val Thr Tyr Ser Leu Gly Lys Phe Gly Gln Arg Ala
            180                 185                 190

Leu Asp Phe Tyr Ser Ile His Val Thr Lys Glu Ser Thr His Pro Val
        195                 200                 205

Lys Pro Leu Ala Gln Ile Ala Gly Asn Arg Tyr Ala Ser Gly Pro Val
    210                 215                 220

Gly Lys Ala Leu Ser Asp Ala Cys Met Gly Thr Ile Ala Ser Phe Leu
225                 230                 235                 240

Ser Lys Tyr Gln Asp Ile Ile Ile Glu His Gln Lys Val Val Lys Gly
                245                 250                 255

Asn Gln Lys Arg Leu Glu Ser Leu Arg Glu Leu Ala Gly Lys Glu Asn
            260                 265                 270
```

```
Leu Glu Tyr Pro Ser Val Thr Leu Pro Pro Gln Pro His Thr Lys Glu
        275                 280                 285

Gly Val Asp Phe Ala Tyr Asn Glu Val Ile Ala Arg Val Arg Met Trp
    290                 295                 300

Val Asn Leu Asn Leu Trp Gln Lys Leu Lys Leu Ser Arg Asp Asp Ala
305                 310                 315                 320

Lys Pro Leu Leu Arg Leu Lys Gly Phe Pro Ser Phe Pro Val Val Glu
                325                 330                 335

Arg Arg Glu Asn Glu Val Asp Trp Trp Asn Thr Ile Asn Glu Val Lys
            340                 345                 350

Lys Leu Ile Asp Ala Lys Arg Asp Met Gly Arg Val Phe Trp Ser Gly
        355                 360                 365

Val Thr Ala Glu Lys Arg Asn Thr Ile Leu Glu Gly Tyr Asn Tyr Leu
    370                 375                 380

Pro Asn Glu Asn Asp His Lys Lys Arg Glu Gly Ser Leu Glu Asn Pro
385                 390                 395                 400

Lys Lys Pro Ala Lys Arg Gln Phe Gly Asp Leu Leu Leu Tyr Leu Glu
                405                 410                 415

Lys Lys Tyr Ala Gly Asp Trp Gly Lys Val Phe Asp Glu Ala Trp Glu
            420                 425                 430

Arg Ile Asp Lys Lys Ile Ala Gly Leu Thr Ser His Ile Glu Arg Glu
        435                 440                 445

Glu Ala Arg Asn Ala Glu Asp Ala Gln Ser Lys Ala Val Leu Thr Asp
    450                 455                 460

Trp Leu Arg Ala Lys Ala Ser Phe Val Leu Glu Arg Leu Lys Glu Met
465                 470                 475                 480
```

```
Asp Glu Lys Glu Phe Tyr Ala Cys Glu Ile Gln Leu Gln Lys Trp Tyr
                485                 490                 495

Gly Asp Leu Arg Gly Asn Pro Phe Ala Val Glu Ala Glu Asn Arg Val
            500                 505                 510

Val Asp Ile Ser Gly Phe Ser Ile Gly Ser Asp Gly His Ser Ile Gln
        515                 520                 525

Tyr Arg Asn Leu Leu Ala Trp Lys Tyr Leu Glu Asn Gly Lys Arg Glu
    530                 535                 540

Phe Tyr Leu Leu Met Asn Tyr Gly Lys Lys Gly Arg Ile Arg Phe Thr
545                 550                 555                 560

Asp Gly Thr Asp Ile Lys Lys Ser Gly Lys Trp Gln Gly Leu Leu Tyr
                565                 570                 575

Gly Gly Gly Lys Ala Lys Val Ile Asp Leu Thr Phe Asp Pro Asp Asp
            580                 585                 590

Glu Gln Leu Ile Ile Leu Pro Leu Ala Phe Gly Thr Arg Gln Gly Arg
        595                 600                 605

Glu Phe Ile Trp Asn Asp Leu Leu Ser Leu Glu Thr Gly Leu Ile Lys
    610                 615                 620

Leu Ala Asn Gly Arg Val Ile Glu Lys Thr Ile Tyr Asn Lys Lys Ile
625                 630                 635                 640

Gly Arg Asp Glu Pro Ala Leu Phe Val Ala Leu Thr Phe Glu Arg Arg
                645                 650                 655

Glu Val Val Asp Pro Ser Asn Ile Lys Pro Val Asn Leu Ile Gly Val
            660                 665                 670

Ala Arg Gly Glu Asn Ile Pro Ala Val Ile Ala Leu Thr Asp Pro Glu
        675                 680                 685

Gly Cys Pro Leu Pro Glu Phe Lys Asp Ser Ser Gly Gly Pro Thr Asp
    690                 695                 700
```

```
Ile Leu Arg Ile Gly Glu Gly Tyr Lys Glu Lys Gln Arg Ala Ile Gln
705                 710                 715                 720

Ala Ala Lys Glu Val Glu Gln Arg Arg Ala Gly Gly Tyr Ser Arg Lys
                725                 730                 735

Phe Ala Ser Lys Ser Arg Asn Leu Ala Asp Asp Met Val Arg Asn Ser
            740                 745                 750

Ala Arg Asp Leu Phe Tyr His Ala Val Thr His Asp Ala Val Leu Val
        755                 760                 765

Phe Ala Asn Leu Ser Arg Gly Phe Gly Arg Gln Gly Lys Arg Thr Phe
    770                 775                 780

Met Thr Glu Arg Gln Tyr Thr Lys Met Glu Asp Trp Leu Thr Ala Lys
785                 790                 795                 800

Leu Ala Tyr Glu Gly Leu Thr Ser Lys Thr Tyr Leu Ser Lys Thr Leu
                805                 810                 815

Ala Gln Tyr Thr Ser Lys Thr Cys Ser Asn Cys Gly Phe Thr Ile Thr
            820                 825                 830

Tyr Ala Asp Met Asp Val Met Leu Val Arg Leu Lys Lys Thr Ser Asp
        835                 840                 845

Gly Trp Ala Thr Thr Leu Asn Asn Lys Glu Leu Lys Ala Glu Tyr Gln
    850                 855                 860

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

Ser Ala Glu Leu Asp Arg Leu Ser Glu Glu Ser Gly Asn Asn Asp Ile
                885                 890                 895

Ser Lys Trp Thr Lys Gly Arg Arg Asp Glu Ala Leu Phe Leu Leu Lys
            900                 905                 910
```

Lys Arg Phe Ser His Arg Pro Val Gln Glu Gln Phe Val Cys Leu Asp
915 920 925

Cys Gly His Glu Val His Ala Ala Glu Gln Ala Ala Leu Asn Ile Ala
930 935 940

Arg Ser Trp Leu Phe Leu Asn Ser Asn Ser Thr Glu Phe Lys Ser Tyr
945 950 955 960

Lys Ser Gly Lys Gln Pro Phe Val Gly Ala Trp Gln Ala Phe Tyr Lys
965 970 975

Arg Arg Leu Lys Glu Val Trp Lys Pro Asn Ala
980 985

<210> 207
<211> 1210
<212> PRT
<213> Parcubacteria group bacterium

<220>
<221> MISC_FEATURE
<222> (1)..(1210)
<223> APG80656.1 CRISPR-associated protein CasY [uncultured Parcubacteria group bacterium]; CasY polypeptide sequence

<400> 207

Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu
1 5 10 15

His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr
20 25 30

Ile Lys Tyr Pro Leu Tyr Ser Ser Pro Ser Gly Gly Arg Thr Val Pro
35 40 45

Arg Glu Ile Val Ser Ala Ile Asn Asp Asp Tyr Val Gly Leu Tyr Gly
50 55 60

Leu Ser Asn Phe Asp Asp Leu Tyr Asn Ala Glu Lys Arg Asn Glu Glu
65 70 75 80

```
Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly
                85                  90                  95

Ala Val Phe Ser Tyr Thr Ala Pro Gly Leu Leu Lys Asn Val Ala Glu
            100                 105                 110

Val Arg Gly Gly Ser Tyr Glu Leu Thr Lys Thr Leu Lys Gly Ser His
        115                 120                 125

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

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

Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln
                165                 170                 175

Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly
            180                 185                 190

Ala Ser Leu Gly Glu Arg Gln Lys Lys Leu Phe Arg Asp Phe Phe Gly
        195                 200                 205

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

Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg
225                 230                 235                 240

Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys
                245                 250                 255

Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile
            260                 265                 270

Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu
        275                 280                 285

Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met
    290                 295                 300
```

```
Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Glu Leu Glu
305                 310                 315                 320

Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro
                325                 330                 335

Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys
            340                 345                 350

Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys
        355                 360                 365

Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met
    370                 375                 380

Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser
385                 390                 395                 400

Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp
                405                 410                 415

Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu
            420                 425                 430

Ser Asp Gly Arg Leu Thr Leu Asn Arg Phe Val Gln Arg Glu Asp Val
        435                 440                 445

Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys
    450                 455                 460

Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr
465                 470                 475                 480

Ile Asp Phe Lys Glu Leu Phe Pro His Leu Ala Lys Pro Leu Lys Leu
                485                 490                 495

Val Pro Asn Phe Tyr Gly Asp Ser Lys Arg Glu Leu Tyr Lys Lys Tyr
            500                 505                 510
```

```
Lys Asn Ala Ala Ile Tyr Thr Asp Ala Leu Trp Lys Ala Val Glu Lys
        515                 520                 525

Ile Tyr Lys Ser Ala Phe Ser Ser Ser Leu Lys Asn Ser Phe Phe Asp
    530                 535                 540

Thr Asp Phe Asp Lys Asp Phe Phe Ile Lys Arg Leu Gln Lys Ile Phe
545                 550                 555                 560

Ser Val Tyr Arg Arg Phe Asn Thr Asp Lys Trp Lys Pro Ile Val Lys
                565                 570                 575

Asn Ser Phe Ala Pro Tyr Cys Asp Ile Val Ser Leu Ala Glu Asn Glu
            580                 585                 590

Val Leu Tyr Lys Pro Lys Gln Ser Arg Ser Arg Lys Ser Ala Ala Ile
        595                 600                 605

Asp Lys Asn Arg Val Arg Leu Pro Ser Thr Glu Asn Ile Ala Lys Ala
    610                 615                 620

Gly Ile Ala Leu Ala Arg Glu Leu Ser Val Ala Gly Phe Asp Trp Lys
625                 630                 635                 640

Asp Leu Leu Lys Lys Glu Glu His Glu Glu Tyr Ile Asp Leu Ile Glu
                645                 650                 655

Leu His Lys Thr Ala Leu Ala Leu Leu Leu Ala Val Thr Glu Thr Gln
            660                 665                 670

Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp
        675                 680                 685

Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu
    690                 695                 700

Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala
705                 710                 715                 720

Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr
                725                 730                 735
```

```
Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln
            740                 745                 750

Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp
        755                 760                 765

Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser
    770                 775                 780

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

Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val
                805                 810                 815

Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu
            820                 825                 830

Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile
        835                 840                 845

Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp
    850                 855                 860

Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly
865                 870                 875                 880

Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp
                885                 890                 895

Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe
            900                 905                 910

Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu
        915                 920                 925

Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr
    930                 935                 940
```

```
Thr Ala Leu Glu Ile Thr Gly Asp Ser Ala Lys Ile Leu Asp Gln Asn
945                 950                 955                 960


Phe Ile Ser Asp Pro Gln Leu Lys Thr Leu Arg Glu Glu Val Lys Gly
                965                 970                 975


Leu Lys Leu Asp Gln Arg Arg Gly Thr Phe Ala Met Pro Ser Thr Lys
            980                 985                 990


Ile Ala Arg Ile Arg Glu Ser Leu  Val His Ser Leu Arg  Asn Arg Ile
        995                 1000                1005


His His  Leu Ala Leu Lys His  Lys Ala Lys Ile Val  Tyr Glu Leu
    1010                1015                1020


Glu Val  Ser Arg Phe Glu Glu  Gly Lys Gln Lys Ile  Lys Lys Val
    1025                1030                1035


Tyr Ala  Thr Leu Lys Lys Ala  Asp Val Tyr Ser Glu  Ile Asp Ala
    1040                1045                1050


Asp Lys  Asn Leu Gln Thr Thr  Val Trp Gly Lys Leu  Ala Val Ala
    1055                1060                1065


Ser Glu  Ile Ser Ala Ser Tyr  Thr Ser Gln Phe Cys  Gly Ala Cys
    1070                1075                1080


Lys Lys  Leu Trp Arg Ala Glu  Met Gln Val Asp Glu  Thr Ile Thr
    1085                1090                1095


Thr Gln  Glu Leu Ile Gly Thr  Val Arg Val Ile Lys  Gly Gly Thr
    1100                1105                1110


Leu Ile  Asp Ala Ile Lys Asp  Phe Met Arg Pro Pro  Ile Phe Asp
    1115                1120                1125


Glu Asn  Asp Thr Pro Phe Pro  Lys Tyr Arg Asp Phe  Cys Asp Lys
    1130                1135                1140


His His  Ile Ser Lys Lys Met  Arg Gly Asn Ser Cys  Leu Phe Ile
    1145                1150                1155
```

Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser
    1160   1165   1170

Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val
    1175   1180   1185

Glu Asp Tyr Phe Glu Arg Phe Arg Lys Leu Lys Asn Ile Lys Val
    1190   1195   1200

Leu Gly Gln Met Lys Lys Ile
    1205   1210

<210> 208
<211> 1082
<212> PRT
<213> Neisseria meningitidis

<220>
<221> MISC_FEATURE
<222> (1)..(1082)
<223> NCBI Reference: WP_002235162.1; type II CRISPR RNA-guided
       endonuclease Cas9; Nme1Cas9 polypeptide sequence

<400> 208

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1   5   10   15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Asp
            20   25   30

Glu Asn Pro Ile Cys Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
        35   40   45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
    50   55   60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65   70   75   80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
                85   90   95

```
Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
            100                 105                 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
        115                 120                 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
    130                 135                 140

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

Gly Val Ala Asp Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
                165                 170                 175

Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
            180                 185                 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300
```

```
Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
                405                 410                 415

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525
```

```
Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
        595                 600                 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
        675                 680                 685

Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
                725                 730                 735
```

```
Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
        755                 760                 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        835                 840                 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
    850                 855                 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865                 870                 875                 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
                885                 890                 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
            900                 905                 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
        915                 920                 925

Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
    930                 935                 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
945                 950                 955                 960
```

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
965 970 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
980 985 990

Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
995 1000 1005

Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys
1010 1015 1020

His Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp
1025 1030 1035

His Lys Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys
1040 1045 1050

Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys
1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1070 1075 1080

<210> 209
<211> 1082
<212> PRT
<213> Neisseria meningitidis

<220>
<221> MISC_FEATURE
<222> (1)..(1082)
<223> NCBI Reference: WP_002230835; type II CRISPR RNA-guided endonuclease Cas9; Nme2Cas9 polypeptide sequence

<400> 209

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

```
Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
        35                  40                  45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
    50                  55                  60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65                  70                  75                  80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
                85                  90                  95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
            100                 105                 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
        115                 120                 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
    130                 135                 140

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

Gly Val Ala Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
                165                 170                 175

Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
            180                 185                 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240
```

```
Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

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

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460
```

```
Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
        595                 600                 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670
```

```
Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr
        675                 680                 685

Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
                725                 730                 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln
        755                 760                 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Ala His Lys Asp Thr Leu Arg Ser
        835                 840                 845

Ala Lys Arg Phe Val Lys His Asn Glu Lys Ile Ser Val Lys Arg Val
    850                 855                 860

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

Lys Asn Gly Arg Glu Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg Leu
                885                 890                 895
```

```
Glu Ala Tyr Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn
            900                 905                 910

Pro Phe Tyr Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val Glu
        915                 920                 925

Lys Thr Gln Glu Ser Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr Thr
    930                 935                 940

Ile Ala Asp Asn Gly Asp Met Val Arg Val Asp Val Phe Cys Lys Val
945                 950                 955                 960

Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile Val Pro Ile Tyr Ala Trp
                965                 970                 975

Gln Val Ala Glu Asn Ile Leu Pro Asp Ile Asp Cys Lys Gly Tyr Arg
            980                 985                 990

Ile Asp Asp Ser Tyr Thr Phe Cys  Phe Ser Leu His Lys  Tyr Asp Leu
        995                 1000                1005

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

Tyr Ile  Asn Cys Asp Ser Ser  Asn Gly Arg Phe Tyr  Leu Ala Trp
    1025                1030                1035

His Asp  Lys Gly Ser Lys Glu  Gln Gln Phe Arg Ile  Ser Thr Gln
    1040                1045                1050

Asn Leu  Val Leu Ile Gln Lys  Tyr Gln Val Asn Glu  Leu Gly Lys
    1055                1060                1065

Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg Pro Pro  Val Arg
    1070                1075                1080
```

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

<220>
<221> MISC_FEATURE
<222> (1)..(1300)
<223> wild type Cpf1 polypeptide sequence

<400> 210

```
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 Asn Val 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>  211
<211>  1300
<212>  PRT
<213>  Francisella novicida


<220>
<221>  MISC_FEATURE
<222>  (1)..(1300)
<223>  Cpf1 D917A polypeptide sequence

<400>  211

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 Asn Val 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 Ala 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>  212
<211>  1300
<212>  PRT
<213>  Francisella novicida


<220>
<221>  MISC_FEATURE
<222>  (1)..(1300)
<223>  Cpf1 E1006A polypeptide sequence

<400>  212

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 Asn Val 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  Ala 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> 213
<211> 1300
<212> PRT
<213> Francisella novicida

<220>
<221> MISC_FEATURE
<222> (1)..(1300)
<223> Cpf1 D1255A polypeptide sequence

<400> 213

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 Asn Val 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 Ala  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>  214
<211>  1300
<212>  PRT
<213>  Francisella novicida


<220>
<221>  MISC_FEATURE
<222>  (1)..(1300)
<223>  Cpf1 D917A/E1006A polypeptide sequence

<400>  214

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 Asn Val 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 Ala 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  Ala 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> 215
<211> 1300
<212> PRT
<213> Francisella novicida

<220>

<221> MISC_FEATURE
<222> (1)..(1300)
<223> Cpf1 D917A/D1255A polypeptide sequence

<400> 215

```
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 Asn Val 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 Ala 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 Ala  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> 216
<211> 1300
<212> PRT
<213> Francisella novicida

<220>
<221> MISC_FEATURE
<222> (1)..(1300)
<223> Cpf1 E1006A/D1255A polypeptide sequence

<400> 216

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 Asn Val 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  Ala 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 Ala  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>  217
<211>  1300
<212>  PRT
<213>  Francisella novicida


<220>
<221>  MISC_FEATURE
<222>  (1)..(1300)
<223>  Cpf1 D917A/E1006A/D1255A polypeptide sequence

<400>  217

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 Asn Val 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 Ala 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  Ala 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 Ala  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> 218
<211> 1052
<212> PRT
<213> Staphylococcus aureus

<220>
<221> MISC_FEATURE
<222> (1)..(1052)
<223> Exemplary SaCas9 polypeptide sequence

<400> 218

Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1 5 10 15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val
20 25 30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser
35 40 45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln
50 55 60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser
65 70 75 80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser
85 90 95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala
100 105 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly
115 120 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu
130 135 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp
145 150 155 160

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

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu
            180                 185                 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
        195                 200                 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
    210                 215                 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225                 230                 235                 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
                245                 250                 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
            260                 265                 270

Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
        275                 280                 285

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
    290                 295                 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305                 310                 315                 320

Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
                325                 330                 335

Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
            340                 345                 350

Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
        355                 360                 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
```

```
    370                 375                 380

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

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile
                405                 410                 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
            420                 425                 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
        435                 440                 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
    450                 455                 460

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

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
                485                 490                 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
            500                 505                 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
        515                 520                 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp
    530                 535                 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545                 550                 555                 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
                565                 570                 575

Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
            580                 585                 590
```

```
Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu
        595                 600                 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr
    610                 615                 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe
625                 630                 635                 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
                645                 650                 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
            660                 665                 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
        675                 680                 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
    690                 695                 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705                 710                 715                 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
                725                 730                 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
            740                 745                 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
        755                 760                 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn
    770                 775                 780

Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785                 790                 795                 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
```

```
                805                 810                 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
            820                 825                 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
        835                 840                 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
    850                 855                 860

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865                 870                 875                 880

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
                885                 890                 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
            900                 905                 910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
        915                 920                 925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
    930                 935                 940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945                 950                 955                 960

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

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
            980                 985                 990

Val Asn Met Ile Asp Ile Thr Tyr  Arg Glu Tyr Leu Glu  Asn Met Asn
        995                 1000                1005

Asp Lys  Arg Pro Pro Arg Ile  Ile Lys Thr Ile Ala  Ser Lys Thr
    1010                1015                1020
```

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
    1025   1030   1035

Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
    1040   1045   1050

<210> 219
<211> 1052
<212> PRT
<213> Staphylococcus aureus

<220>
<221> MISC_FEATURE
<222> (1)..(1052)
<223> SaCas9n polypeptide sequence

<400> 219

Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1   5   10   15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val
            20   25   30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser
        35   40   45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln
    50   55   60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser
65   70   75   80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser
                85   90   95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala
            100   105   110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly
        115   120   125

```
Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu
    130                 135                 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp
145                 150                 155                 160

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

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu
            180                 185                 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
        195                 200                 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
    210                 215                 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225                 230                 235                 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
                245                 250                 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
            260                 265                 270

Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
        275                 280                 285

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
    290                 295                 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305                 310                 315                 320

Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
                325                 330                 335

Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
            340                 345                 350
```

```
Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
        355                 360                 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
    370                 375                 380

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

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile
                405                 410                 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
            420                 425                 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
        435                 440                 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
    450                 455                 460

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

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
                485                 490                 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
            500                 505                 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
        515                 520                 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp
    530                 535                 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545                 550                 555                 560
```

```
Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
                565                 570                 575

Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
            580                 585                 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu
        595                 600                 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr
    610                 615                 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe
625                 630                 635                 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
                645                 650                 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
            660                 665                 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
        675                 680                 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
    690                 695                 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705                 710                 715                 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
                725                 730                 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
            740                 745                 750

Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
        755                 760                 765

Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn
    770                 775                 780
```

```
Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785                 790                 795                 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
                805                 810                 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
            820                 825                 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
        835                 840                 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
    850                 855                 860

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865                 870                 875                 880

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
                885                 890                 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
            900                 905                 910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
        915                 920                 925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
    930                 935                 940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945                 950                 955                 960

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

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
            980                 985                 990
```

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
        995   1000   1005

Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr
    1010   1015   1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
    1025   1030   1035

Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
    1040   1045   1050

<210> 220
<211> 1052
<212> PRT
<213> Staphylococcus aureus

<220>
<221> MISC_FEATURE
<222> (1)..(1052)
<223> SaKKH Cas9 polypeptide sequence

<400> 220

Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly
1   5   10   15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val
            20   25   30

Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg Ser
        35   40   45

Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile Gln
    50   55   60

Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His Ser
65   70   75   80

Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu Ser
                85   90   95

Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu Ala

```
            100                 105                 110

Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr Gly
        115                 120                 125

Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala Leu
    130                 135                 140

Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys Asp
145                 150                 155                 160

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

Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln Leu
            180                 185                 190

Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg Arg
        195                 200                 205

Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys Asp
    210                 215                 220

Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe Pro
225                 230                 235                 240

Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr Asn
                245                 250                 255

Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn Glu
            260                 265                 270

Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe Lys
        275                 280                 285

Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu Val
    290                 295                 300

Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys Pro
305                 310                 315                 320
```

```
Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr Ala
                325                 330                 335

Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala Lys
            340                 345                 350

Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu Thr
        355                 360                 365

Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser Asn
    370                 375                 380

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

Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala Ile
                405                 410                 415

Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln Gln
            420                 425                 430

Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro Val
        435                 440                 445

Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile Ile
    450                 455                 460

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

Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys Arg
                485                 490                 495

Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr Gly
            500                 505                 510

Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp Met
        515                 520                 525

Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu Asp
```

```
    530                 535                 540

Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro Arg
545                 550                 555                 560

Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys Gln
                565                 570                 575

Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu Ser
            580                 585                 590

Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile Leu
        595                 600                 605

Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu Tyr
    610                 615                 620

Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp Phe
625                 630                 635                 640

Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu Met
                645                 650                 655

Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys Val
            660                 665                 670

Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp Lys
        675                 680                 685

Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp Ala
    690                 695                 700

Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys Leu
705                 710                 715                 720

Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
                725                 730                 735

Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
            740                 745                 750
```

```
Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
        755                 760                 765

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

Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile
785                 790                 795                 800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys
                805                 810                 815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp
            820                 825                 830

Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp
        835                 840                 845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu
    850                 855                 860

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys
865                 870                 875                 880

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr
                885                 890                 895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg
            900                 905                 910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys
        915                 920                 925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys
    930                 935                 940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu
945                 950                 955                 960

Phe Ile Ala Ser Phe Tyr Lys Asn Asp Leu Ile Lys Ile Asn Gly Glu
```

965 970 975

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu
980 985 990

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn
995 1000 1005

Asp Lys Arg Pro Pro His Ile Ile Lys Thr Ile Ala Ser Lys Thr
1010 1015 1020

Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr
1025 1030 1035

Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 221
<211> 707
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-1 polypeptide sequence

<400> 221

Met Ala Asp Thr Pro Thr Leu Phe Thr Gln Phe Leu Arg His His Leu
1 5 10 15

Pro Gly Gln Arg Phe Arg Lys Asp Ile Leu Lys Gln Ala Gly Arg Ile
20 25 30

Leu Ala Asn Lys Gly Glu Asp Ala Thr Ile Ala Phe Leu Arg Gly Lys
35 40 45

Ser Glu Glu Ser Pro Pro Asp Phe Gln Pro Pro Val Lys Cys Pro Ile
50 55 60

Ile Ala Cys Ser Arg Pro Leu Thr Glu Trp Pro Ile Tyr Gln Ala Ser
65 70 75 80

Val Ala Ile Gln Gly Tyr Val Tyr Gly Gln Ser Leu Ala Glu Phe Glu

```
                85                  90                  95

Ala Ser Asp Pro Gly Cys Ser Lys Asp Gly Leu Leu Gly Trp Phe Asp
            100                 105                 110

Lys Thr Gly Val Cys Thr Asp Tyr Phe Ser Val Gln Gly Leu Asn Leu
        115                 120                 125

Ile Phe Gln Asn Ala Arg Lys Arg Tyr Ile Gly Val Gln Thr Lys Val
    130                 135                 140

Thr Asn Arg Asn Glu Lys Arg His Lys Lys Leu Lys Arg Ile Asn Ala
145                 150                 155                 160

Lys Arg Ile Ala Glu Gly Leu Pro Glu Leu Thr Ser Asp Glu Pro Glu
                165                 170                 175

Ser Ala Leu Asp Glu Thr Gly His Leu Ile Asp Pro Pro Gly Leu Asn
            180                 185                 190

Thr Asn Ile Tyr Cys Tyr Gln Gln Val Ser Pro Lys Pro Leu Ala Leu
        195                 200                 205

Ser Glu Val Asn Gln Leu Pro Thr Ala Tyr Ala Gly Tyr Ser Thr Ser
    210                 215                 220

Gly Asp Asp Pro Ile Gln Pro Met Val Thr Lys Asp Arg Leu Ser Ile
225                 230                 235                 240

Ser Lys Gly Gln Pro Gly Tyr Ile Pro Glu His Gln Arg Ala Leu Leu
                245                 250                 255

Ser Gln Lys Lys His Arg Arg Met Arg Gly Tyr Gly Leu Lys Ala Arg
            260                 265                 270

Ala Leu Leu Val Ile Val Arg Ile Gln Asp Asp Trp Ala Val Ile Asp
        275                 280                 285

Leu Arg Ser Leu Leu Arg Asn Ala Tyr Trp Arg Arg Ile Val Gln Thr
    290                 295                 300
```

```
Lys Glu Pro Ser Thr Ile Thr Lys Leu Leu Lys Leu Val Thr Gly Asp
305                 310                 315                 320

Pro Val Leu Asp Ala Thr Arg Met Val Ala Thr Phe Thr Tyr Lys Pro
                325                 330                 335

Gly Ile Val Gln Val Arg Ser Ala Lys Cys Leu Lys Asn Lys Gln Gly
            340                 345                 350

Ser Lys Leu Phe Ser Glu Arg Tyr Leu Asn Glu Thr Val Ser Val Thr
        355                 360                 365

Ser Ile Asp Leu Gly Ser Asn Asn Leu Val Ala Val Ala Thr Tyr Arg
    370                 375                 380

Leu Val Asn Gly Asn Thr Pro Glu Leu Leu Gln Arg Phe Thr Leu Pro
385                 390                 395                 400

Ser His Leu Val Lys Asp Phe Glu Arg Tyr Lys Gln Ala His Asp Thr
                405                 410                 415

Leu Glu Asp Ser Ile Gln Lys Thr Ala Val Ala Ser Leu Pro Gln Gly
            420                 425                 430

Gln Gln Thr Glu Ile Arg Met Trp Ser Met Tyr Gly Phe Arg Glu Ala
        435                 440                 445

Gln Glu Arg Val Cys Gln Glu Leu Gly Leu Ala Asp Gly Ser Ile Pro
    450                 455                 460

Trp Asn Val Met Thr Ala Thr Ser Thr Ile Leu Thr Asp Leu Phe Leu
465                 470                 475                 480

Ala Arg Gly Gly Asp Pro Lys Lys Cys Met Phe Thr Ser Glu Pro Lys
                485                 490                 495

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

Trp Ala Lys Met Tyr Arg Thr Leu Leu Ser Lys Glu Thr Arg Glu Ala
```

```
        515                 520                 525

Trp Asn Lys Ala Leu Trp Gly Leu Lys Arg Gly Ser Pro Asp Tyr Ala
    530                 535                 540

Arg Leu Ser Lys Arg Lys Glu Glu Leu Ala Arg Arg Cys Val Asn Tyr
545                 550                 555                 560

Thr Ile Ser Thr Ala Glu Lys Arg Ala Gln Cys Gly Arg Thr Ile Val
                565                 570                 575

Ala Leu Glu Asp Leu Asn Ile Gly Phe Phe His Gly Arg Gly Lys Gln
            580                 585                 590

Glu Pro Gly Trp Val Gly Leu Phe Thr Arg Lys Lys Glu Asn Arg Trp
        595                 600                 605

Leu Met Gln Ala Leu His Lys Ala Phe Leu Glu Leu Ala His His Arg
    610                 615                 620

Gly Tyr His Val Ile Glu Val Asn Pro Ala Tyr Thr Ser Gln Thr Cys
625                 630                 635                 640

Pro Val Cys Arg His Cys Asp Pro Asp Asn Arg Asp Gln His Asn Arg
                645                 650                 655

Glu Ala Phe His Cys Ile Gly Cys Gly Phe Arg Gly Asn Ala Asp Leu
            660                 665                 670

Asp Val Ala Thr His Asn Ile Ala Met Val Ala Ile Thr Gly Glu Ser
        675                 680                 685

Leu Lys Arg Ala Arg Gly Ser Val Ala Ser Lys Thr Pro Gln Pro Leu
    690                 695                 700

Ala Ala Glu
705
```

<210> 222
<211> 757
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-2 polypeptide sequence

<400> 222

Met Pro Lys Pro Ala Val Glu Ser Glu Phe Ser Lys Val Leu Lys Lys
1 5 10 15

His Phe Pro Gly Glu Arg Phe Arg Ser Ser Tyr Met Lys Arg Gly Gly
20 25 30

Lys Ile Leu Ala Ala Gln Gly Glu Glu Ala Val Val Ala Tyr Leu Gln
35 40 45

Gly Lys Ser Glu Glu Glu Pro Pro Asn Phe Gln Pro Pro Ala Lys Cys
50 55 60

His Val Val Thr Lys Ser Arg Asp Phe Ala Glu Trp Pro Ile Met Lys
65 70 75 80

Ala Ser Glu Ala Ile Gln Arg Tyr Ile Tyr Ala Leu Ser Thr Thr Glu
85 90 95

Arg Ala Ala Cys Lys Pro Gly Lys Ser Ser Glu Ser His Ala Ala Trp
100 105 110

Phe Ala Ala Thr Gly Val Ser Asn His Gly Tyr Ser His Val Gln Gly
115 120 125

Leu Asn Leu Ile Phe Asp His Thr Leu Gly Arg Tyr Asp Gly Val Leu
130 135 140

Lys Lys Val Gln Leu Arg Asn Glu Lys Ala Arg Ala Arg Leu Glu Ser
145 150 155 160

Ile Asn Ala Ser Arg Ala Asp Glu Gly Leu Pro Glu Ile Lys Ala Glu
165 170 175

Glu Glu Glu Val Ala Thr Asn Glu Thr Gly His Leu Leu Gln Pro Pro
180 185 190

```
Gly Ile Asn Pro Ser Phe Tyr Val Tyr Gln Thr Ile Ser Pro Gln Ala
        195                 200                 205

Tyr Arg Pro Arg Asp Glu Ile Val Leu Pro Pro Glu Tyr Ala Gly Tyr
    210                 215                 220

Val Arg Asp Pro Asn Ala Pro Ile Pro Leu Gly Val Val Arg Asn Arg
225                 230                 235                 240

Cys Asp Ile Gln Lys Gly Cys Pro Gly Tyr Ile Pro Glu Trp Gln Arg
                245                 250                 255

Glu Ala Gly Thr Ala Ile Ser Pro Lys Thr Gly Lys Ala Val Thr Val
            260                 265                 270

Pro Gly Leu Ser Pro Lys Lys Asn Lys Arg Met Arg Arg Tyr Trp Arg
        275                 280                 285

Ser Glu Lys Glu Lys Ala Gln Asp Ala Leu Leu Val Thr Val Arg Ile
    290                 295                 300

Gly Thr Asp Trp Val Val Ile Asp Val Arg Gly Leu Leu Arg Asn Ala
305                 310                 315                 320

Arg Trp Arg Thr Ile Ala Pro Lys Asp Ile Ser Leu Asn Ala Leu Leu
                325                 330                 335

Asp Leu Phe Thr Gly Asp Pro Val Ile Asp Val Arg Arg Asn Ile Val
            340                 345                 350

Thr Phe Thr Tyr Thr Leu Asp Ala Cys Gly Thr Tyr Ala Arg Lys Trp
        355                 360                 365

Thr Leu Lys Gly Lys Gln Thr Lys Ala Thr Leu Asp Lys Leu Thr Ala
    370                 375                 380

Thr Gln Thr Val Ala Leu Val Ala Ile Asp Leu Gly Gln Thr Asn Pro
385                 390                 395                 400

Ile Ser Ala Gly Ile Ser Arg Val Thr Gln Glu Asn Gly Ala Leu Gln
```

```
               405                 410                 415

Cys Glu Pro Leu Asp Arg Phe Thr Leu Pro Asp Asp Leu Leu Lys Asp
            420                 425                 430

Ile Ser Ala Tyr Arg Ile Ala Trp Asp Arg Asn Glu Glu Glu Leu Arg
        435                 440                 445

Ala Arg Ser Val Glu Ala Leu Pro Glu Ala Gln Gln Ala Glu Val Arg
    450                 455                 460

Ala Leu Asp Gly Val Ser Lys Glu Thr Ala Arg Thr Gln Leu Cys Ala
465                 470                 475                 480

Asp Phe Gly Leu Asp Pro Lys Arg Leu Pro Trp Asp Lys Met Ser Ser
                485                 490                 495

Asn Thr Thr Phe Ile Ser Glu Ala Leu Leu Ser Asn Ser Val Ser Arg
            500                 505                 510

Asp Gln Val Phe Phe Thr Pro Ala Pro Lys Lys Gly Ala Lys Lys Lys
        515                 520                 525

Ala Pro Val Glu Val Met Arg Lys Asp Arg Thr Trp Ala Arg Ala Tyr
    530                 535                 540

Lys Pro Arg Leu Ser Val Glu Ala Gln Lys Leu Lys Asn Glu Ala Leu
545                 550                 555                 560

Trp Ala Leu Lys Arg Thr Ser Pro Glu Tyr Leu Lys Leu Ser Arg Arg
                565                 570                 575

Lys Glu Glu Leu Cys Arg Arg Ser Ile Asn Tyr Val Ile Glu Lys Thr
            580                 585                 590

Arg Arg Arg Thr Gln Cys Gln Ile Val Ile Pro Val Ile Glu Asp Leu
        595                 600                 605

Asn Val Arg Phe Phe His Gly Ser Gly Lys Arg Leu Pro Gly Trp Asp
    610                 615                 620
```

```
Asn Phe Phe Thr Ala Lys Lys Glu Asn Arg Trp Phe Ile Gln Gly Leu
625                 630                 635                 640


His Lys Ala Phe Ser Asp Leu Arg Thr His Arg Ser Phe Tyr Val Phe
                645                 650                 655


Glu Val Arg Pro Glu Arg Thr Ser Ile Thr Cys Pro Lys Cys Gly His
            660                 665                 670


Cys Glu Val Gly Asn Arg Asp Gly Glu Ala Phe Gln Cys Leu Ser Cys
        675                 680                 685


Gly Lys Thr Cys Asn Ala Asp Leu Asp Val Ala Thr His Asn Leu Thr
    690                 695                 700


Gln Val Ala Leu Thr Gly Lys Thr Met Pro Lys Arg Glu Glu Pro Arg
705                 710                 715                 720


Asp Ala Gln Gly Thr Ala Pro Ala Arg Lys Thr Lys Lys Ala Ser Lys
                725                 730                 735


Ser Lys Ala Pro Pro Ala Glu Arg Glu Asp Gln Thr Pro Ala Gln Glu
            740                 745                 750


Pro Ser Gln Thr Ser
        755


<210>  223
<211>  766
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: Casphi-3 polypeptide sequence

<400>  223

Met Glu Lys Glu Ile Thr Glu Leu Thr Lys Ile Arg Arg Glu Phe Pro
1               5                   10                  15


Asn Lys Lys Phe Ser Ser Thr Asp Met Lys Lys Ala Gly Lys Leu Leu
            20                  25                  30
```

```
Lys Ala Glu Gly Pro Asp Ala Val Arg Asp Phe Leu Asn Ser Cys Gln
        35                  40                  45

Glu Ile Ile Gly Asp Phe Lys Pro Pro Val Lys Thr Asn Ile Val Ser
    50                  55                  60

Ile Ser Arg Pro Phe Glu Glu Trp Pro Val Ser Met Val Gly Arg Ala
65                  70                  75                  80

Ile Gln Glu Tyr Tyr Phe Ser Leu Thr Lys Glu Glu Leu Glu Ser Val
                85                  90                  95

His Pro Gly Thr Ser Ser Glu Asp His Lys Ser Phe Phe Asn Ile Thr
            100                 105                 110

Gly Leu Ser Asn Tyr Asn Tyr Thr Ser Val Gln Gly Leu Asn Leu Ile
        115                 120                 125

Phe Lys Asn Ala Lys Ala Ile Tyr Asp Gly Thr Leu Val Lys Ala Asn
    130                 135                 140

Asn Lys Asn Lys Lys Leu Glu Lys Lys Phe Asn Glu Ile Asn His Lys
145                 150                 155                 160

Arg Ser Leu Glu Gly Leu Pro Ile Ile Thr Pro Asp Phe Glu Glu Pro
                165                 170                 175

Phe Asp Glu Asn Gly His Leu Asn Asn Pro Pro Gly Ile Asn Arg Asn
            180                 185                 190

Ile Tyr Gly Tyr Gln Gly Cys Ala Ala Lys Val Phe Val Pro Ser Lys
        195                 200                 205

His Lys Met Val Ser Leu Pro Lys Glu Tyr Glu Gly Tyr Asn Arg Asp
    210                 215                 220

Pro Asn Leu Ser Leu Ala Gly Phe Arg Asn Arg Leu Glu Ile Pro Glu
225                 230                 235                 240

Gly Glu Pro Gly His Val Pro Trp Phe Gln Arg Met Asp Ile Pro Glu
```

```
               245                 250                 255

Gly Gln Ile Gly His Val Asn Lys Ile Gln Arg Phe Asn Phe Val His
            260                 265                 270

Gly Lys Asn Ser Gly Lys Val Lys Phe Ser Asp Lys Thr Gly Arg Val
        275                 280                 285

Lys Arg Tyr His His Ser Lys Tyr Lys Asp Ala Thr Lys Pro Tyr Lys
    290                 295                 300

Phe Leu Glu Glu Ser Lys Lys Val Ser Ala Leu Asp Ser Ile Leu Ala
305                 310                 315                 320

Ile Ile Thr Ile Gly Asp Asp Trp Val Val Phe Asp Ile Arg Gly Leu
               325                 330                 335

Tyr Arg Asn Val Phe Tyr Arg Glu Leu Ala Gln Lys Gly Leu Thr Ala
            340                 345                 350

Val Gln Leu Leu Asp Leu Phe Thr Gly Asp Pro Val Ile Asp Pro Lys
        355                 360                 365

Lys Gly Val Val Thr Phe Ser Tyr Lys Glu Gly Val Val Pro Val Phe
    370                 375                 380

Ser Gln Lys Ile Val Pro Arg Phe Lys Ser Arg Asp Thr Leu Glu Lys
385                 390                 395                 400

Leu Thr Ser Gln Gly Pro Val Ala Leu Leu Ser Val Asp Leu Gly Gln
               405                 410                 415

Asn Glu Pro Val Ala Ala Arg Val Cys Ser Leu Lys Asn Ile Asn Asp
            420                 425                 430

Lys Ile Thr Leu Asp Asn Ser Cys Arg Ile Ser Phe Leu Asp Asp Tyr
        435                 440                 445

Lys Lys Gln Ile Lys Asp Tyr Arg Asp Ser Leu Asp Glu Leu Glu Ile
    450                 455                 460
```

```
Lys Ile Arg Leu Glu Ala Ile Asn Ser Leu Glu Thr Asn Gln Gln Val
465                 470                 475                 480


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


Thr Val Asp Met Phe Asp Ile Asp Pro Asn Leu Ile Ser Trp Asp Ser
            500                 505                 510


Met Ser Asp Ala Arg Val Ser Thr Gln Ile Ser Asp Leu Tyr Leu Lys
        515                 520                 525


Asn Gly Gly Asp Glu Ser Arg Val Tyr Phe Glu Ile Asn Asn Lys Arg
    530                 535                 540


Ile Lys Arg Ser Asp Tyr Asn Ile Ser Gln Leu Val Arg Pro Lys Leu
545                 550                 555                 560


Ser Asp Ser Thr Arg Lys Asn Leu Asn Asp Ser Ile Trp Lys Leu Lys
                565                 570                 575


Arg Thr Ser Glu Glu Tyr Leu Lys Leu Ser Lys Arg Lys Leu Glu Leu
            580                 585                 590


Ser Arg Ala Val Val Asn Tyr Thr Ile Arg Gln Ser Lys Leu Leu Ser
        595                 600                 605


Gly Ile Asn Asp Ile Val Ile Ile Leu Glu Asp Leu Asp Val Lys Lys
    610                 615                 620


Lys Phe Asn Gly Arg Gly Ile Arg Asp Ile Gly Trp Asp Asn Phe Phe
625                 630                 635                 640


Ser Ser Arg Lys Glu Asn Arg Trp Phe Ile Pro Ala Phe His Lys Ala
                645                 650                 655


Phe Ser Glu Leu Ser Ser Asn Arg Gly Leu Cys Val Ile Glu Val Asn
            660                 665                 670


Pro Ala Trp Thr Ser Ala Thr Cys Pro Asp Cys Gly Phe Cys Ser Lys
```

675 680 685

Glu Asn Arg Asp Gly Ile Asn Phe Thr Cys Arg Lys Cys Gly Val Ser
690 695 700

Tyr His Ala Asp Ile Asp Val Ala Thr Leu Asn Ile Ala Arg Val Ala
705 710 715 720

Val Leu Gly Lys Pro Met Ser Gly Pro Ala Asp Arg Glu Arg Leu Gly
725 730 735

Asp Thr Lys Lys Pro Arg Val Ala Arg Ser Arg Lys Thr Met Lys Arg
740 745 750

Lys Asp Ile Ser Asn Ser Thr Val Glu Ala Met Val Thr Ala
755 760 765

<210> 224
<211> 765
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-4 polypeptide sequence

<400> 224

Met Tyr Ser Leu Glu Met Ala Asp Leu Lys Ser Glu Pro Ser Leu Leu
1 5 10 15

Ala Lys Leu Leu Arg Asp Arg Phe Pro Gly Lys Tyr Trp Leu Pro Lys
20 25 30

Tyr Trp Lys Leu Ala Glu Lys Lys Arg Leu Thr Gly Gly Glu Glu Ala
35 40 45

Ala Cys Glu Tyr Met Ala Asp Lys Gln Leu Asp Ser Pro Pro Pro Asn
50 55 60

Phe Arg Pro Pro Ala Arg Cys Val Ile Leu Ala Lys Ser Arg Pro Phe
65 70 75 80

Glu Asp Trp Pro Val His Arg Val Ala Ser Lys Ala Gln Ser Phe Val

```
                85                  90                  95

Ile Gly Leu Ser Glu Gln Gly Phe Ala Ala Leu Arg Ala Ala Pro Pro
            100                 105                 110

Ser Thr Ala Asp Ala Arg Arg Asp Trp Leu Arg Ser His Gly Ala Ser
        115                 120                 125

Glu Asp Asp Leu Met Ala Leu Glu Ala Gln Leu Leu Glu Thr Ile Met
    130                 135                 140

Gly Asn Ala Ile Ser Leu His Gly Gly Val Leu Lys Lys Ile Asp Asn
145                 150                 155                 160

Ala Asn Val Lys Ala Ala Lys Arg Leu Ser Gly Arg Asn Glu Ala Arg
                165                 170                 175

Leu Asn Lys Gly Leu Gln Glu Leu Pro Pro Glu Gln Glu Gly Ser Ala
            180                 185                 190

Tyr Gly Ala Asp Gly Leu Leu Val Asn Pro Pro Gly Leu Asn Leu Asn
        195                 200                 205

Ile Tyr Cys Arg Lys Ser Cys Cys Pro Lys Pro Val Lys Asn Thr Ala
    210                 215                 220

Arg Phe Val Gly His Tyr Pro Gly Tyr Leu Arg Asp Ser Asp Ser Ile
225                 230                 235                 240

Leu Ile Ser Gly Thr Met Asp Arg Leu Thr Ile Ile Glu Gly Met Pro
                245                 250                 255

Gly His Ile Pro Ala Trp Gln Arg Glu Gln Gly Leu Val Lys Pro Gly
            260                 265                 270

Gly Arg Arg Arg Arg Leu Ser Gly Ser Glu Ser Asn Met Arg Gln Lys
        275                 280                 285

Val Asp Pro Ser Thr Gly Pro Arg Arg Ser Thr Arg Ser Gly Thr Val
    290                 295                 300
```

```
Asn Arg Ser Asn Gln Arg Thr Gly Arg Asn Gly Asp Pro Leu Leu Val
305                 310                 315                 320

Glu Ile Arg Met Lys Glu Asp Trp Val Leu Leu Asp Ala Arg Gly Leu
                325                 330                 335

Leu Arg Asn Leu Arg Trp Arg Glu Ser Lys Arg Gly Leu Ser Cys Asp
            340                 345                 350

His Glu Asp Leu Ser Leu Ser Gly Leu Leu Ala Leu Phe Ser Gly Asp
        355                 360                 365

Pro Val Ile Asp Pro Val Arg Asn Glu Val Val Phe Leu Tyr Gly Glu
    370                 375                 380

Gly Ile Ile Pro Val Arg Ser Thr Lys Pro Val Gly Thr Arg Gln Ser
385                 390                 395                 400

Lys Lys Leu Leu Glu Arg Gln Ala Ser Met Gly Pro Leu Thr Leu Ile
                405                 410                 415

Ser Cys Asp Leu Gly Gln Thr Asn Leu Ile Ala Gly Arg Ala Ser Ala
            420                 425                 430

Ile Ser Leu Thr His Gly Ser Leu Gly Val Arg Ser Ser Val Arg Ile
        435                 440                 445

Glu Leu Asp Pro Glu Ile Ile Lys Ser Phe Glu Arg Leu Arg Lys Asp
    450                 455                 460

Ala Asp Arg Leu Glu Thr Glu Ile Leu Thr Ala Ala Lys Glu Thr Leu
465                 470                 475                 480

Ser Asp Glu Gln Arg Gly Glu Val Asn Ser His Glu Lys Asp Ser Pro
                485                 490                 495

Gln Thr Ala Lys Ala Ser Leu Cys Arg Glu Leu Gly Leu His Pro Pro
            500                 505                 510

Ser Leu Pro Trp Gly Gln Met Gly Pro Ser Thr Thr Phe Ile Ala Asp
```

```
        515                 520                 525

Met Leu Ile Ser His Gly Arg Asp Asp Asp Ala Phe Leu Ser His Gly
    530                 535                 540

Glu Phe Pro Thr Leu Glu Lys Arg Lys Lys Phe Asp Lys Arg Phe Cys
545                 550                 555                 560

Leu Glu Ser Arg Pro Leu Leu Ser Ser Glu Thr Arg Lys Ala Leu Asn
                565                 570                 575

Glu Ser Leu Trp Glu Val Lys Arg Thr Ser Ser Glu Tyr Ala Arg Leu
            580                 585                 590

Ser Gln Arg Lys Lys Glu Met Ala Arg Arg Ala Val Asn Phe Val Val
        595                 600                 605

Glu Ile Ser Arg Arg Lys Thr Gly Leu Ser Asn Val Ile Val Asn Ile
    610                 615                 620

Glu Asp Leu Asn Val Arg Ile Phe His Gly Gly Gly Lys Gln Ala Pro
625                 630                 635                 640

Gly Trp Asp Gly Phe Phe Arg Pro Lys Ser Glu Asn Arg Trp Phe Ile
                645                 650                 655

Gln Ala Ile His Lys Ala Phe Ser Asp Leu Ala Ala His His Gly Ile
            660                 665                 670

Pro Val Ile Glu Ser Asp Pro Gln Arg Thr Ser Met Thr Cys Pro Glu
        675                 680                 685

Cys Gly His Cys Asp Ser Lys Asn Arg Asn Gly Val Arg Phe Leu Cys
    690                 695                 700

Lys Gly Cys Gly Ala Ser Met Asp Ala Asp Phe Asp Ala Ala Cys Arg
705                 710                 715                 720

Asn Leu Glu Arg Val Ala Leu Thr Gly Lys Pro Met Pro Lys Pro Ser
                725                 730                 735
```

Thr Ser Cys Glu Arg Leu Leu Ser Ala Thr Thr Gly Lys Val Cys Ser
740 745 750

Asp His Ser Leu Ser His Asp Ala Ile Glu Lys Ala Ser
755 760 765

<210> 225
<211> 793
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-5 polypeptide sequence

<400> 225

Met Ser Ser Leu Pro Thr Pro Leu Glu Leu Leu Lys Gln Lys His Ala
1 5 10 15

Asp Leu Phe Lys Gly Leu Gln Phe Ser Ser Lys Asp Asn Lys Met Ala
20 25 30

Gly Lys Val Leu Lys Lys Asp Gly Glu Glu Ala Ala Leu Ala Phe Leu
35 40 45

Ser Glu Arg Gly Val Ser Arg Gly Glu Leu Pro Asn Phe Arg Pro Pro
50 55 60

Ala Lys Thr Leu Val Val Ala Gln Ser Arg Pro Phe Glu Glu Phe Pro
65 70 75 80

Ile Tyr Arg Val Ser Glu Ala Ile Gln Leu Tyr Val Tyr Ser Leu Ser
85 90 95

Val Lys Glu Leu Glu Thr Val Pro Ser Gly Ser Ser Thr Lys Lys Glu
100 105 110

His Gln Arg Phe Phe Gln Asp Ser Ser Val Pro Asp Phe Gly Tyr Thr
115 120 125

Ser Val Gln Gly Leu Asn Lys Ile Phe Gly Leu Ala Arg Gly Ile Tyr
130 135 140

```
Leu Gly Val Ile Thr Arg Gly Glu Asn Gln Leu Gln Lys Ala Lys Ser
145                 150                 155                 160

Lys His Glu Ala Leu Asn Lys Lys Arg Arg Ala Ser Gly Glu Ala Glu
                165                 170                 175

Thr Glu Phe Asp Pro Thr Pro Tyr Glu Tyr Met Thr Pro Glu Arg Lys
            180                 185                 190

Leu Ala Lys Pro Pro Gly Val Asn His Ser Ile Met Cys Tyr Val Asp
        195                 200                 205

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

Pro Ser Glu Tyr Ala Gly Tyr Cys Arg Glu Ile Asn Thr Ala Ile Glu
225                 230                 235                 240

Lys Gly Thr Val Asp Arg Leu Gly His Leu Lys Gly Gly Pro Gly Tyr
                245                 250                 255

Ile Pro Gly His Gln Arg Lys Glu Ser Thr Thr Glu Gly Pro Lys Ile
            260                 265                 270

Asn Phe Arg Lys Gly Arg Ile Arg Arg Ser Tyr Thr Ala Leu Tyr Ala
        275                 280                 285

Lys Arg Asp Ser Arg Arg Val Arg Gln Gly Lys Leu Ala Leu Pro Ser
    290                 295                 300

Tyr Arg His His Met Met Arg Leu Asn Ser Asn Ala Glu Ser Ala Ile
305                 310                 315                 320

Leu Ala Val Ile Phe Phe Gly Lys Asp Trp Val Val Phe Asp Leu Arg
                325                 330                 335

Gly Leu Leu Arg Asn Val Arg Trp Arg Asn Leu Phe Val Asp Gly Ser
            340                 345                 350

Thr Pro Ser Thr Leu Leu Gly Met Phe Gly Asp Pro Val Ile Asp Pro
```

```
        355                 360                 365

Lys Arg Gly Val Val Ala Phe Cys Tyr Lys Glu Gln Ile Val Pro Val
    370                 375                 380

Val Ser Lys Ser Ile Thr Lys Met Val Lys Ala Pro Glu Leu Leu Asn
385                 390                 395                 400

Lys Leu Tyr Leu Lys Ser Glu Asp Pro Leu Val Leu Val Ala Ile Asp
                405                 410                 415

Leu Gly Gln Thr Asn Pro Val Gly Val Gly Val Tyr Arg Val Met Asn
            420                 425                 430

Ala Ser Leu Asp Tyr Glu Val Val Thr Arg Phe Ala Leu Glu Ser Glu
        435                 440                 445

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

Ala Gln Ile Arg Ala Glu Thr Phe Asp Ala Met Thr Ser Glu Glu Gln
465                 470                 475                 480

Glu Glu Ile Thr Arg Val Arg Ala Phe Ser Ala Ser Lys Ala Lys Glu
                485                 490                 495

Asn Val Cys His Arg Phe Gly Met Pro Val Asp Ala Val Asp Trp Ala
            500                 505                 510

Thr Met Gly Ser Asn Thr Ile His Ile Ala Lys Trp Val Met Arg His
        515                 520                 525

Gly Asp Pro Ser Leu Val Glu Val Leu Glu Tyr Arg Lys Asp Asn Glu
    530                 535                 540

Ile Lys Leu Asp Lys Asn Gly Val Pro Lys Lys Val Lys Leu Thr Asp
545                 550                 555                 560

Lys Arg Ile Ala Asn Leu Thr Ser Ile Arg Leu Arg Phe Ser Gln Glu
                565                 570                 575
```

```
Thr Ser Lys His Tyr Asn Asp Thr Met Trp Glu Leu Arg Arg Lys His
            580                 585                 590

Pro Val Tyr Gln Lys Leu Ser Lys Ser Lys Ala Asp Phe Ser Arg Arg
        595                 600                 605

Val Val Asn Ser Ile Ile Arg Arg Val Asn His Leu Val Pro Arg Ala
    610                 615                 620

Arg Ile Val Phe Ile Ile Glu Asp Leu Lys Asn Leu Gly Lys Val Phe
625                 630                 635                 640

His Gly Ser Gly Lys Arg Glu Leu Gly Trp Asp Ser Tyr Phe Glu Pro
                645                 650                 655

Lys Ser Glu Asn Arg Trp Phe Ile Gln Val Leu His Lys Ala Phe Ser
            660                 665                 670

Glu Thr Gly Lys His Lys Gly Tyr Tyr Ile Ile Glu Cys Trp Pro Asn
        675                 680                 685

Trp Thr Ser Cys Thr Cys Pro Lys Cys Ser Cys Cys Asp Ser Glu Asn
    690                 695                 700

Arg His Gly Glu Val Phe Arg Cys Leu Ala Cys Gly Tyr Thr Cys Asn
705                 710                 715                 720

Thr Asp Phe Gly Thr Ala Pro Asp Asn Leu Val Lys Ile Ala Thr Thr
                725                 730                 735

Gly Lys Gly Leu Pro Gly Pro Lys Lys Arg Cys Lys Gly Ser Ser Lys
            740                 745                 750

Gly Lys Asn Pro Lys Ile Ala Arg Ser Ser Glu Thr Gly Val Ser Val
        755                 760                 765

Thr Glu Ser Gly Ala Pro Lys Val Lys Lys Ser Ser Pro Thr Gln Thr
    770                 775                 780

Ser Gln Ser Ser Ser Gln Ser Ala Pro
```

785 790

<210> 226
<211> 441
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-6 polypeptide sequence

<400> 226

Met Asn Lys Ile Glu Lys Glu Lys Thr Pro Leu Ala Lys Leu Met Asn
1 5 10 15

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

Gly Lys Lys Leu Leu Lys Glu Gly Glu Leu Lys Thr Ile Glu Tyr Met
35 40 45

Thr Gly Lys Gly Ser Ile Glu Pro Leu Pro Asn Phe Lys Pro Pro Val
50 55 60

Lys Cys Leu Ile Val Ala Lys Arg Arg Asp Leu Lys Tyr Phe Pro Ile
65 70 75 80

Cys Lys Ala Ser Cys Glu Ile Gln Ser Tyr Val Tyr Ser Leu Asn Tyr
85 90 95

Lys Asp Phe Met Asp Tyr Phe Ser Thr Pro Met Thr Ser Gln Lys Gln
100 105 110

His Glu Glu Phe Phe Lys Lys Ser Gly Leu Asn Ile Glu Tyr Gln Asn
115 120 125

Val Ala Gly Leu Asn Leu Ile Phe Asn Asn Val Lys Asn Thr Tyr Asn
130 135 140

Gly Val Ile Leu Lys Val Lys Asn Arg Asn Glu Lys Leu Lys Lys Lys
145 150 155 160

Ala Ile Lys Asn Asn Tyr Glu Phe Glu Glu Ile Lys Thr Phe Asn Asp

```
                165                 170                 175

Asp Gly Cys Leu Ile Asn Lys Pro Gly Ile Asn Asn Val Ile Tyr Cys
            180                 185                 190

Phe Gln Ser Ile Ser Pro Lys Ile Leu Lys Asn Ile Thr His Leu Pro
        195                 200                 205

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

Lys Tyr Val Ser Arg Leu Asp Ile Pro Glu Ser Gln Pro Gly His Val
225                 230                 235                 240

Pro Glu Trp Gln Arg Lys Leu Pro Glu Phe Asn Asn Thr Asn Asn Pro
                245                 250                 255

Arg Arg Arg Arg Lys Trp Tyr Ser Asn Gly Arg Asn Ile Ser Lys Gly
            260                 265                 270

Tyr Ser Val Asp Gln Val Asn Gln Ala Lys Ile Glu Asp Ser Leu Leu
        275                 280                 285

Ala Gln Ile Lys Ile Gly Glu Asp Trp Ile Ile Leu Asp Ile Arg Gly
    290                 295                 300

Leu Leu Arg Asp Leu Asn Arg Arg Glu Leu Ile Ser Tyr Lys Asn Lys
305                 310                 315                 320

Leu Thr Ile Lys Asp Val Leu Gly Phe Phe Ser Asp Tyr Pro Ile Ile
                325                 330                 335

Asp Ile Lys Lys Asn Leu Val Thr Phe Cys Tyr Lys Glu Gly Val Ile
            340                 345                 350

Gln Val Val Ser Gln Lys Ser Ile Gly Asn Lys Lys Ser Lys Gln Leu
        355                 360                 365

Leu Glu Lys Leu Ile Glu Asn Lys Pro Ile Ala Leu Val Ser Ile Asp
    370                 375                 380
```

Leu Gly Gln Thr Asn Pro Val Ser Val Lys Ile Ser Lys Leu Asn Lys
385 390 395 400

Ile Asn Asn Lys Ile Ser Ile Glu Ser Phe Thr Tyr Arg Phe Leu Asn
405 410 415

Glu Glu Ile Leu Lys Glu Ile Glu Lys Tyr Arg Lys Asp Tyr Asp Lys
420 425 430

Leu Glu Leu Lys Leu Ile Asn Glu Ala
435 440

<210> 227
<211> 772
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-7 polypeptide sequence

<400> 227

Met Ser Asn Thr Ala Val Ser Thr Arg Glu His Met Ser Asn Lys Thr
1 5 10 15

Thr Pro Pro Ser Pro Leu Ser Leu Leu Leu Arg Ala His Phe Pro Gly
20 25 30

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

Asp Gly Gly Pro Glu Ala Val Ile Ser Tyr Leu Thr Gly Lys Gly Gln
50 55 60

Ala Lys Leu Lys Asp Val Lys Pro Pro Ala Lys Ala Phe Val Ile Ala
65 70 75 80

Gln Ser Arg Pro Phe Ile Glu Trp Asp Leu Val Arg Val Ser Arg Gln
85 90 95

Ile Gln Glu Lys Ile Phe Gly Ile Pro Ala Thr Lys Gly Arg Pro Lys
100 105 110

```
Gln Asp Gly Leu Ser Glu Thr Ala Phe Asn Glu Ala Val Ala Ser Leu
        115                 120                 125

Glu Val Asp Gly Lys Ser Lys Leu Asn Glu Glu Thr Arg Ala Ala Phe
    130                 135                 140

Tyr Glu Val Leu Gly Leu Asp Ala Pro Ser Leu His Ala Gln Ala Gln
145                 150                 155                 160

Asn Ala Leu Ile Lys Ser Ala Ile Ser Ile Arg Glu Gly Val Leu Lys
                165                 170                 175

Lys Val Glu Asn Arg Asn Glu Lys Asn Leu Ser Lys Thr Lys Arg Arg
            180                 185                 190

Lys Glu Ala Gly Glu Glu Ala Thr Phe Val Glu Glu Lys Ala His Asp
        195                 200                 205

Glu Arg Gly Tyr Leu Ile His Pro Pro Gly Val Asn Gln Thr Ile Pro
    210                 215                 220

Gly Tyr Gln Ala Val Val Ile Lys Ser Cys Pro Ser Asp Phe Ile Gly
225                 230                 235                 240

Leu Pro Ser Gly Cys Leu Ala Lys Glu Ser Ala Glu Ala Leu Thr Asp
                245                 250                 255

Tyr Leu Pro His Asp Arg Met Thr Ile Pro Lys Gly Gln Pro Gly Tyr
            260                 265                 270

Val Pro Glu Trp Gln His Pro Leu Leu Asn Arg Arg Lys Asn Arg Arg
        275                 280                 285

Arg Arg Asp Trp Tyr Ser Ala Ser Leu Asn Lys Pro Lys Ala Thr Cys
    290                 295                 300

Ser Lys Arg Ser Gly Thr Pro Asn Arg Lys Asn Ser Arg Thr Asp Gln
305                 310                 315                 320

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

```
                325                 330                 335

Gln Asp Glu Trp Val Ile Ile Asp Ile Arg Gly Leu Leu Arg Asn Ala
            340                 345                 350

Arg Tyr Arg Lys Leu Leu Lys Glu Lys Ser Thr Ile Pro Asp Leu Leu
        355                 360                 365

Ser Leu Phe Thr Gly Asp Pro Ser Ile Asp Met Arg Gln Gly Val Cys
    370                 375                 380

Thr Phe Ile Tyr Lys Ala Gly Gln Ala Cys Ser Ala Lys Met Val Lys
385                 390                 395                 400

Thr Lys Asn Ala Pro Glu Ile Leu Ser Glu Leu Thr Lys Ser Gly Pro
                405                 410                 415

Val Val Leu Val Ser Ile Asp Leu Gly Gln Thr Asn Pro Ile Ala Ala
            420                 425                 430

Lys Val Ser Arg Val Thr Gln Leu Ser Asp Gly Gln Leu Ser His Glu
        435                 440                 445

Thr Leu Leu Arg Glu Leu Leu Ser Asn Asp Ser Ser Asp Gly Lys Glu
    450                 455                 460

Ile Ala Arg Tyr Arg Val Ala Ser Asp Arg Leu Arg Asp Lys Leu Ala
465                 470                 475                 480

Asn Leu Ala Val Glu Arg Leu Ser Pro Glu His Lys Ser Glu Ile Leu
                485                 490                 495

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

Ala Leu Gly Leu Asn Pro Glu Met Ile Ala Trp Asp Lys Met Thr Pro
        515                 520                 525

Tyr Thr Glu Phe Leu Ala Thr Ala Tyr Leu Glu Lys Gly Gly Asp Arg
    530                 535                 540
```

```
Lys Val Ala Thr Leu Lys Pro Lys Asn Arg Pro Glu Met Leu Arg Arg
545                 550                 555                 560


Asp Ile Lys Phe Lys Gly Thr Glu Gly Val Arg Ile Glu Val Ser Pro
                565                 570                 575


Glu Ala Ala Glu Ala Tyr Arg Glu Ala Gln Trp Asp Leu Gln Arg Thr
            580                 585                 590


Ser Pro Glu Tyr Leu Arg Leu Ser Thr Trp Lys Gln Glu Leu Thr Lys
        595                 600                 605


Arg Ile Leu Asn Gln Leu Arg His Lys Ala Ala Lys Ser Ser Gln Cys
    610                 615                 620


Glu Val Val Val Met Ala Phe Glu Asp Leu Asn Ile Lys Met Met His
625                 630                 635                 640


Gly Asn Gly Lys Trp Ala Asp Gly Gly Trp Asp Ala Phe Phe Ile Lys
                645                 650                 655


Lys Arg Glu Asn Arg Trp Phe Met Gln Ala Phe His Lys Ser Leu Thr
            660                 665                 670


Glu Leu Gly Ala His Lys Gly Val Pro Thr Ile Glu Val Thr Pro His
        675                 680                 685


Arg Thr Ser Ile Thr Cys Thr Lys Cys Gly His Cys Asp Lys Ala Asn
    690                 695                 700


Arg Asp Gly Glu Arg Phe Ala Cys Gln Lys Cys Gly Phe Val Ala His
705                 710                 715                 720


Ala Asp Leu Glu Ile Ala Thr Asp Asn Ile Glu Arg Val Ala Leu Thr
                725                 730                 735


Gly Lys Pro Met Pro Lys Pro Glu Ser Glu Arg Ser Gly Asp Ala Lys
            740                 745                 750


Lys Ser Val Gly Ala Arg Lys Ala Ala Phe Lys Pro Glu Glu Asp Ala
```

```
        755                 760                 765


Glu Ala Ala Glu
    770


<210>  228
<211>  717
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: Casphi-8 polypeptide sequence

<400>  228

Met Ile Lys Pro Thr Val Ser Gln Phe Leu Thr Pro Gly Phe Lys Leu
1               5                   10                  15


Ile Arg Asn His Ser Arg Thr Ala Gly Leu Lys Leu Lys Asn Glu Gly
            20                  25                  30


Glu Glu Ala Cys Lys Lys Phe Val Arg Glu Asn Glu Ile Pro Lys Asp
        35                  40                  45


Glu Cys Pro Asn Phe Gln Gly Gly Pro Ala Ile Ala Asn Ile Ile Ala
    50                  55                  60


Lys Ser Arg Glu Phe Thr Glu Trp Glu Ile Tyr Gln Ser Ser Leu Ala
65                  70                  75                  80


Ile Gln Glu Val Ile Phe Thr Leu Pro Lys Asp Lys Leu Pro Glu Pro
                85                  90                  95


Ile Leu Lys Glu Glu Trp Arg Ala Gln Trp Leu Ser Glu His Gly Leu
            100                 105                 110


Asp Thr Val Pro Tyr Lys Glu Ala Ala Gly Leu Asn Leu Ile Ile Lys
        115                 120                 125


Asn Ala Val Asn Thr Tyr Lys Gly Val Gln Val Lys Val Asp Asn Lys
    130                 135                 140


Asn Lys Asn Asn Leu Ala Lys Ile Asn Arg Lys Asn Glu Ile Ala Lys
```

```
145                 150                 155                 160

Leu Asn Gly Glu Gln Glu Ile Ser Phe Glu Glu Ile Lys Ala Phe Asp
                165                 170                 175

Asp Lys Gly Tyr Leu Leu Gln Lys Pro Ser Pro Asn Lys Ser Ile Tyr
            180                 185                 190

Cys Tyr Gln Ser Val Ser Pro Lys Pro Phe Ile Thr Ser Lys Tyr His
        195                 200                 205

Asn Val Asn Leu Pro Glu Glu Tyr Ile Gly Tyr Tyr Arg Lys Ser Asn
    210                 215                 220

Glu Pro Ile Val Ser Pro Tyr Gln Phe Asp Arg Leu Arg Ile Pro Ile
225                 230                 235                 240

Gly Glu Pro Gly Tyr Val Pro Lys Trp Gln Tyr Thr Phe Leu Ser Lys
                245                 250                 255

Lys Glu Asn Lys Arg Arg Lys Leu Ser Lys Arg Ile Lys Asn Val Ser
            260                 265                 270

Pro Ile Leu Gly Ile Ile Cys Ile Lys Lys Asp Trp Cys Val Phe Asp
        275                 280                 285

Met Arg Gly Leu Leu Arg Thr Asn His Trp Lys Lys Tyr His Lys Pro
    290                 295                 300

Thr Asp Ser Ile Asn Asp Leu Phe Asp Tyr Phe Thr Gly Asp Pro Val
305                 310                 315                 320

Ile Asp Thr Lys Ala Asn Val Val Arg Phe Arg Tyr Lys Met Glu Asn
                325                 330                 335

Gly Ile Val Asn Tyr Lys Pro Val Arg Glu Lys Lys Gly Lys Glu Leu
            340                 345                 350

Leu Glu Asn Ile Cys Asp Gln Asn Gly Ser Cys Lys Leu Ala Thr Val
        355                 360                 365
```

```
Asp Val Gly Gln Asn Asn Pro Val Ala Ile Gly Leu Phe Glu Leu Lys
    370                 375                 380

Lys Val Asn Gly Glu Leu Thr Lys Thr Leu Ile Ser Arg His Pro Thr
385                 390                 395                 400

Pro Ile Asp Phe Cys Asn Lys Ile Thr Ala Tyr Arg Glu Arg Tyr Asp
                405                 410                 415

Lys Leu Glu Ser Ser Ile Lys Leu Asp Ala Ile Lys Gln Leu Thr Ser
            420                 425                 430

Glu Gln Lys Ile Glu Val Asp Asn Tyr Asn Asn Asn Phe Thr Pro Gln
        435                 440                 445

Asn Thr Lys Gln Ile Val Cys Ser Lys Leu Asn Ile Asn Pro Asn Asp
    450                 455                 460

Leu Pro Trp Asp Lys Met Ile Ser Gly Thr His Phe Ile Ser Glu Lys
465                 470                 475                 480

Ala Gln Val Ser Asn Lys Ser Glu Ile Tyr Phe Thr Ser Thr Asp Lys
                485                 490                 495

Gly Lys Thr Lys Asp Val Met Lys Ser Asp Tyr Lys Trp Phe Gln Asp
            500                 505                 510

Tyr Lys Pro Lys Leu Ser Lys Glu Val Arg Asp Ala Leu Ser Asp Ile
        515                 520                 525

Glu Trp Arg Leu Arg Arg Glu Ser Leu Glu Phe Asn Lys Leu Ser Lys
    530                 535                 540

Ser Arg Glu Gln Asp Ala Arg Gln Leu Ala Asn Trp Ile Ser Ser Met
545                 550                 555                 560

Cys Asp Val Ile Gly Ile Glu Asn Leu Val Lys Lys Asn Asn Phe Phe
                565                 570                 575

Gly Gly Ser Gly Lys Arg Glu Pro Gly Trp Asp Asn Phe Tyr Lys Pro
```

580 585 590

Lys Lys Glu Asn Arg Trp Trp Ile Asn Ala Ile His Lys Ala Leu Thr
595 600 605

Glu Leu Ser Gln Asn Lys Gly Lys Arg Val Ile Leu Leu Pro Ala Met
610 615 620

Arg Thr Ser Ile Thr Cys Pro Lys Cys Lys Tyr Cys Asp Ser Lys Asn
625 630 635 640

Arg Asn Gly Glu Lys Phe Asn Cys Leu Lys Cys Gly Ile Glu Leu Asn
645 650 655

Ala Asp Ile Asp Val Ala Thr Glu Asn Leu Ala Thr Val Ala Ile Thr
660 665 670

Ala Gln Ser Met Pro Lys Pro Thr Cys Glu Arg Ser Gly Asp Ala Lys
675 680 685

Lys Pro Val Arg Ala Arg Lys Ala Lys Ala Pro Glu Phe His Asp Lys
690 695 700

Leu Ala Pro Ser Tyr Thr Val Val Leu Arg Glu Ala Val
705 710 715

<210> 229
<211> 793
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-9 polypeptide sequence

<400> 229

Met Arg Ser Ser Arg Glu Ile Gly Asp Lys Ile Leu Met Arg Gln Pro
1 5 10 15

Ala Glu Lys Thr Ala Phe Gln Val Phe Arg Gln Glu Val Ile Gly Thr
20 25 30

Gln Lys Leu Ser Gly Gly Asp Ala Lys Thr Ala Gly Arg Leu Tyr Lys

```
        35                  40                  45

Gln Gly Lys Met Glu Ala Ala Arg Glu Trp Leu Leu Lys Gly Ala Arg
    50                  55                  60

Asp Asp Val Pro Pro Asn Phe Gln Pro Pro Ala Lys Cys Leu Val Val
65                  70                  75                  80

Ala Val Ser His Pro Phe Glu Glu Trp Asp Ile Ser Lys Thr Asn His
                85                  90                  95

Asp Val Gln Ala Tyr Ile Tyr Ala Gln Pro Leu Gln Ala Glu Gly His
            100                 105                 110

Leu Asn Gly Leu Ser Glu Lys Trp Glu Asp Thr Ser Ala Asp Gln His
        115                 120                 125

Lys Leu Trp Phe Glu Lys Thr Gly Val Pro Asp Arg Gly Leu Pro Val
    130                 135                 140

Gln Ala Ile Asn Lys Ile Ala Lys Ala Ala Val Asn Arg Ala Phe Gly
145                 150                 155                 160

Val Val Arg Lys Val Glu Asn Arg Asn Glu Lys Arg Arg Ser Arg Asp
                165                 170                 175

Asn Arg Ile Ala Glu His Asn Arg Glu Asn Gly Leu Thr Glu Val Val
            180                 185                 190

Arg Glu Ala Pro Glu Val Ala Thr Asn Ala Asp Gly Phe Leu Leu His
        195                 200                 205

Pro Pro Gly Ile Asp Pro Ser Ile Leu Ser Tyr Ala Ser Val Ser Pro
    210                 215                 220

Val Pro Tyr Asn Ser Ser Lys His Ser Phe Val Arg Leu Pro Glu Glu
225                 230                 235                 240

Tyr Gln Ala Tyr Asn Val Glu Pro Asp Ala Pro Ile Pro Gln Phe Val
                245                 250                 255
```

```
Val Glu Asp Arg Phe Ala Ile Pro Pro Gly Gln Pro Gly Tyr Val Pro
            260                 265                 270

Glu Trp Gln Arg Leu Lys Cys Ser Thr Asn Lys His Arg Arg Met Arg
        275                 280                 285

Gln Trp Ser Asn Gln Asp Tyr Lys Pro Lys Ala Gly Arg Arg Ala Lys
    290                 295                 300

Pro Leu Glu Phe Gln Ala His Leu Thr Arg Glu Arg Ala Lys Gly Ala
305                 310                 315                 320

Leu Leu Val Val Met Arg Ile Lys Glu Asp Trp Val Val Phe Asp Val
                325                 330                 335

Arg Gly Leu Leu Arg Asn Val Glu Trp Arg Lys Val Leu Ser Glu Glu
            340                 345                 350

Ala Arg Glu Lys Leu Thr Leu Lys Gly Leu Leu Asp Leu Phe Thr Gly
        355                 360                 365

Asp Pro Val Ile Asp Thr Lys Arg Gly Ile Val Thr Phe Leu Tyr Lys
    370                 375                 380

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

Ala Arg Asp Leu Leu Leu Arg Leu Thr Glu Pro Gly Glu Asp Gly Leu
                405                 410                 415

Arg Arg Glu Val Gly Leu Val Ala Val Asp Leu Gly Gln Thr His Pro
            420                 425                 430

Ile Ala Ala Ala Ile Tyr Arg Ile Gly Arg Thr Ser Ala Gly Ala Leu
        435                 440                 445

Glu Ser Thr Val Leu His Arg Gln Gly Leu Arg Glu Asp Gln Lys Glu
    450                 455                 460

Lys Leu Lys Glu Tyr Arg Lys Arg His Thr Ala Leu Asp Ser Arg Leu
```

```
465                 470                 475                 480

Arg Lys Glu Ala Phe Glu Thr Leu Ser Val Glu Gln Gln Lys Glu Ile
                485                 490                 495

Val Thr Val Ser Gly Ser Gly Ala Gln Ile Thr Lys Asp Lys Val Cys
            500                 505                 510

Asn Tyr Leu Gly Val Asp Pro Ser Thr Leu Pro Trp Glu Lys Met Gly
        515                 520                 525

Ser Tyr Thr His Phe Ile Ser Asp Asp Phe Leu Arg Arg Gly Gly Asp
    530                 535                 540

Pro Asn Ile Val His Phe Asp Arg Gln Pro Lys Lys Gly Lys Val Ser
545                 550                 555                 560

Lys Lys Ser Gln Arg Ile Lys Arg Ser Asp Ser Gln Trp Val Gly Arg
                565                 570                 575

Met Arg Pro Arg Leu Ser Gln Glu Thr Ala Lys Ala Arg Met Glu Ala
            580                 585                 590

Asp Trp Ala Ala Gln Asn Glu Asn Glu Glu Tyr Lys Arg Leu Ala Arg
        595                 600                 605

Ser Lys Gln Glu Leu Ala Arg Trp Cys Val Asn Thr Leu Leu Gln Asn
    610                 615                 620

Thr Arg Cys Ile Thr Gln Cys Asp Glu Ile Val Val Val Ile Glu Asp
625                 630                 635                 640

Leu Asn Val Lys Ser Leu His Gly Lys Gly Ala Arg Glu Pro Gly Trp
                645                 650                 655

Asp Asn Phe Phe Thr Pro Lys Thr Glu Asn Arg Trp Phe Ile Gln Ile
            660                 665                 670

Leu His Lys Thr Phe Ser Glu Leu Pro Lys His Arg Gly Glu His Val
        675                 680                 685
```

```
Ile Glu Gly Cys Pro Leu Arg Thr Ser Ile Thr Cys Pro Ala Cys Ser
    690                 695                 700

Tyr Cys Asp Lys Asn Ser Arg Asn Gly Glu Lys Phe Val Cys Val Ala
705                 710                 715                 720

Cys Gly Ala Thr Phe His Ala Asp Phe Glu Val Ala Thr Tyr Asn Leu
                725                 730                 735

Val Arg Leu Ala Thr Thr Gly Met Pro Met Pro Lys Ser Leu Glu Arg
            740                 745                 750

Gln Gly Gly Gly Glu Lys Ala Gly Gly Ala Arg Lys Ala Arg Lys Lys
        755                 760                 765

Ala Lys Gln Val Glu Lys Ile Val Val Gln Ala Asn Ala Asn Val Thr
    770                 775                 780

Met Asn Gly Ala Ser Leu His Ser Pro
785                 790
```

<210> 230
<211> 812
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Casphi-10 polypeptide sequence

<400> 230

```
Met Asp Met Leu Asp Thr Glu Thr Asn Tyr Ala Thr Glu Thr Pro Ala
1               5                   10                  15

Gln Gln Gln Asp Tyr Ser Pro Lys Pro Pro Lys Lys Ala Gln Arg Ala
            20                  25                  30

Pro Lys Gly Phe Ser Lys Lys Ala Arg Pro Glu Lys Lys Pro Pro Lys
        35                  40                  45

Pro Ile Thr Leu Phe Thr Gln Lys His Phe Ser Gly Val Arg Phe Leu
    50                  55                  60
```

```
Lys Arg Val Ile Arg Asp Ala Ser Lys Ile Leu Lys Leu Ser Glu Ser
65                  70                  75                  80

Arg Thr Ile Thr Phe Leu Glu Gln Ala Ile Glu Arg Asp Gly Ser Ala
                85                  90                  95

Pro Pro Asp Val Thr Pro Pro Val His Asn Thr Ile Met Ala Val Thr
            100                 105                 110

Arg Pro Phe Glu Glu Trp Pro Glu Val Ile Leu Ser Lys Ala Leu Gln
        115                 120                 125

Lys His Cys Tyr Ala Leu Thr Lys Lys Ile Lys Ile Lys Thr Trp Pro
    130                 135                 140

Lys Lys Gly Pro Gly Lys Lys Cys Leu Ala Ala Trp Ser Ala Arg Thr
145                 150                 155                 160

Lys Ile Pro Leu Ile Pro Gly Gln Val Gln Ala Thr Asn Gly Leu Phe
                165                 170                 175

Asp Arg Ile Gly Ser Ile Tyr Asp Gly Val Glu Lys Lys Val Thr Asn
            180                 185                 190

Arg Asn Ala Asn Lys Lys Leu Glu Tyr Asp Glu Ala Ile Lys Glu Gly
        195                 200                 205

Arg Asn Pro Ala Val Pro Glu Tyr Glu Thr Ala Tyr Asn Ile Asp Gly
    210                 215                 220

Thr Leu Ile Asn Lys Pro Gly Tyr Asn Pro Asn Leu Tyr Ile Thr Gln
225                 230                 235                 240

Ser Arg Thr Pro Arg Leu Ile Thr Glu Ala Asp Arg Pro Leu Val Glu
                245                 250                 255

Lys Ile Leu Trp Gln Met Val Glu Lys Lys Thr Gln Ser Arg Asn Gln
            260                 265                 270
```

```
Ala Arg Arg Ala Arg Leu Glu Lys Ala Ala His Leu Gln Gly Leu Pro
        275                 280                 285

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

Ile Arg Ile Ile Asp Pro Leu Asp Lys Ile Glu Pro Tyr Met Pro Gln
305                 310                 315                 320

Asp Arg Met Ala Ile Lys Ala Ser Gln Asp Gly His Val Pro Tyr Trp
                325                 330                 335

Gln Arg Pro Phe Leu Ser Lys Arg Arg Asn Arg Arg Val Arg Ala Gly
            340                 345                 350

Trp Gly Lys Gln Val Ser Ser Ile Gln Ala Trp Leu Thr Gly Ala Leu
        355                 360                 365

Leu Val Ile Val Arg Leu Gly Asn Glu Ala Phe Leu Ala Asp Ile Arg
    370                 375                 380

Gly Ala Leu Arg Asn Ala Gln Trp Arg Lys Leu Leu Lys Pro Asp Ala
385                 390                 395                 400

Thr Tyr Gln Ser Leu Phe Asn Leu Phe Thr Gly Asp Pro Val Val Asn
                405                 410                 415

Thr Arg Thr Asn His Leu Thr Met Ala Tyr Arg Glu Gly Val Val Asn
            420                 425                 430

Ile Val Lys Ser Arg Ser Phe Lys Gly Arg Gln Thr Arg Glu His Leu
        435                 440                 445

Leu Thr Leu Leu Gly Gln Gly Lys Thr Val Ala Gly Val Ser Phe Asp
    450                 455                 460

Leu Gly Gln Lys His Ala Ala Gly Leu Leu Ala Ala His Phe Gly Leu
465                 470                 475                 480

Gly Glu Asp Gly Asn Pro Val Phe Thr Pro Ile Gln Ala Cys Phe Leu
                485                 490                 495
```

```
Pro Gln Arg Tyr Leu Asp Ser Leu Thr Asn Tyr Arg Asn Arg Tyr Asp
            500                 505                 510

Ala Leu Thr Leu Asp Met Arg Arg Gln Ser Leu Leu Ala Leu Thr Pro
        515                 520                 525

Ala Gln Gln Gln Glu Phe Ala Asp Ala Gln Arg Asp Pro Gly Gly Gln
    530                 535                 540

Ala Lys Arg Ala Cys Cys Leu Lys Leu Asn Leu Asn Pro Asp Glu Ile
545                 550                 555                 560

Arg Trp Asp Leu Val Ser Gly Ile Ser Thr Met Ile Ser Asp Leu Tyr
                565                 570                 575

Ile Glu Arg Gly Gly Asp Pro Arg Asp Val His Gln Gln Val Glu Thr
            580                 585                 590

Lys Pro Lys Gly Lys Arg Lys Ser Glu Ile Arg Ile Leu Lys Ile Arg
        595                 600                 605

Asp Gly Lys Trp Ala Tyr Asp Phe Arg Pro Lys Ile Ala Asp Glu Thr
    610                 615                 620

Arg Lys Ala Gln Arg Glu Gln Leu Trp Lys Leu Gln Lys Ala Ser Ser
625                 630                 635                 640

Glu Phe Glu Arg Leu Ser Arg Tyr Lys Ile Asn Ile Ala Arg Ala Ile
                645                 650                 655

Ala Asn Trp Ala Leu Gln Trp Gly Arg Glu Leu Ser Gly Cys Asp Ile
            660                 665                 670

Val Ile Pro Val Leu Glu Asp Leu Asn Val Gly Ser Lys Phe Phe Asp
        675                 680                 685

Gly Lys Gly Lys Trp Leu Leu Gly Trp Asp Asn Arg Phe Thr Pro Lys
    690                 695                 700
```

```
Lys Glu Asn Arg Trp Phe Ile Lys Val Leu His Lys Ala Val Ala Glu
705                 710                 715                 720


Leu Ala Pro His Arg Gly Val Pro Val Tyr Glu Val Met Pro His Arg
                725                 730                 735


Thr Ser Met Thr Cys Pro Ala Cys His Tyr Cys His Pro Thr Asn Arg
            740                 745                 750


Glu Gly Asp Arg Phe Glu Cys Gln Ser Cys His Val Val Lys Asn Thr
        755                 760                 765


Asp Arg Asp Val Ala Pro Tyr Asn Ile Leu Arg Val Ala Val Glu Gly
    770                 775                 780


Lys Thr Leu Asp Arg Trp Gln Ala Glu Lys Lys Pro Gln Ala Glu Pro
785                 790                 795                 800


Asp Arg Pro Met Ile Leu Ile Asp Asn Gln Glu Ser
                805                 810
```

<210> 231
<211> 84
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 231

```
Met Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu
1               5                   10                  15


Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val
            20                  25                  30


Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp
        35                  40                  45


Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu
    50                  55                  60
```

```
Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys
65                  70                  75                  80


Ile Lys Met Leu



<210>  232
<211>  1129
<212>  PRT
<213>  Alicyclobacillus acido-terrestris


<220>
<221>  MISC_FEATURE
<222>  (1)..(1129)
<223>  Cas12b/C2c1; sp|T0D7A2|C2C1_ALIAG CRISPR-associated endonuclease
       C2c1 OS = Alicyclobacillus acido-terrestris (strain ATCC 49025 /
       DSM 3922/ CIP 106132 / NCIMB 13137/GD3B) GN=c2c1 PE=1 SV=1;
       Cas12b/C2c1 polypeptide sequence

<400>  232

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Asp Asp Met Pro
1               5                   10                  15


Glu Ile Arg Ala Gly Leu Trp Lys Leu His Lys Glu Val Asn Ala Gly
            20                  25                  30


Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu
        35                  40                  45


Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Asp Lys Thr
    50                  55                  60


Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65                  70                  75                  80


Val Glu Asn Gly His Arg Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu
                85                  90                  95


Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
            100                 105                 110


Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu
```

```
        115                 120                 125

Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn
    130                 135                 140

Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu
145                 150                 155                 160

Glu Glu Lys Glu Lys Ala Glu Thr Arg Lys Ser Ala Asp Arg Thr Ala
                165                 170                 175

Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg
            180                 185                 190

Val Tyr Thr Asp Ser Glu Met Ser Ser Val Glu Trp Lys Pro Leu Arg
        195                 200                 205

Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala
    210                 215                 220

Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Gln
225                 230                 235                 240

Glu Tyr Ala Lys Leu Val Glu Gln Lys Asn Arg Phe Glu Gln Lys Asn
                245                 250                 255

Phe Val Gly Gln Glu His Leu Val His Leu Val Asn Gln Leu Gln Gln
            260                 265                 270

Asp Met Lys Glu Ala Ser Pro Gly Leu Glu Ser Lys Glu Gln Thr Ala
        275                 280                 285

His Tyr Val Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu
    290                 295                 300

Lys Trp Gly Lys Leu Ala Pro Asp Ala Pro Phe Asp Leu Tyr Asp Ala
305                 310                 315                 320

Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His
                325                 330                 335
```

```
Asp Leu Phe Ala Lys Leu Ala Glu Pro Glu Tyr Gln Ala Leu Trp Arg
            340                 345                 350

Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Leu
        355                 360                 365

Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
    370                 375                 380

Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
385                 390                 395                 400

Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Arg Arg His
                405                 410                 415

Ala Ile Arg Phe His Lys Leu Leu Lys Val Glu Asn Gly Val Ala Arg
            420                 425                 430

Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Glu Gln Leu Asp
        435                 440                 445

Asn Leu Leu Pro Arg Asp Pro Asn Glu Pro Ile Ala Leu Tyr Phe Arg
    450                 455                 460

Asp Tyr Gly Ala Glu Gln His Phe Thr Gly Glu Phe Gly Gly Ala Lys
465                 470                 475                 480

Ile Gln Cys Arg Arg Asp Gln Leu Ala His Met His Arg Arg Arg Gly
                485                 490                 495

Ala Arg Asp Val Tyr Leu Asn Val Ser Val Arg Val Gln Ser Gln Ser
            500                 505                 510

Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
        515                 520                 525

Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp
    530                 535                 540

Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu
```

```
545                 550                 555                 560

Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser
                565                 570                 575

Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
            580                 585                 590

Asn Ser Lys Gly Arg Val Pro Phe Phe Phe Pro Ile Lys Gly Asn Asp
        595                 600                 605

Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly
    610                 615                 620

Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg
625                 630                 635                 640

Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val
                645                 650                 655

Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys
            660                 665                 670

Leu Ile Glu Gln Pro Val Asp Ala Ala Asn His Met Thr Pro Asp Trp
        675                 680                 685

Arg Glu Ala Phe Glu Asn Glu Leu Gln Lys Leu Lys Ser Leu His Gly
    690                 695                 700

Ile Cys Ser Asp Lys Glu Trp Met Asp Ala Val Tyr Glu Ser Val Arg
705                 710                 715                 720

Arg Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp
                725                 730                 735

Val Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Ala Lys Asp Val
            740                 745                 750

Val Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr
        755                 760                 765
```

```
Lys Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val
    770                 775                 780

Ile Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His
785                 790                 795                 800

Ile Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile
                805                 810                 815

Ile Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Glu Arg Gly Lys
            820                 825                 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu
        835                 840                 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn
    850                 855                 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Ile
865                 870                 875                 880

Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala
                885                 890                 895

Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys
            900                 905                 910

Arg Arg Val Pro Ala Arg Cys Thr Gln Glu His Asn Pro Glu Pro Phe
        915                 920                 925

Pro Trp Trp Leu Asn Lys Phe Val Val Glu His Thr Leu Asp Ala Cys
    930                 935                 940

Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Ile Phe
945                 950                 955                 960

Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala
                965                 970                 975

Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg Leu Trp Ser Asp Phe
```

980 985 990

Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly
995 1000 1005

Glu Leu Val Leu Ile Pro Arg Leu Thr Gly Lys Arg Thr Ala Asp
1010 1015 1020

Ser Tyr Ser Asn Lys Val Phe Tyr Thr Asn Thr Gly Val Thr Tyr
1025 1030 1035

Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln
1040 1045 1050

Glu Lys Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp
1055 1060 1065

Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly
1070 1075 1080

Ile Ile Asn Arg Gly Asn Trp Thr Arg Gln Lys Glu Phe Trp Ser
1085 1090 1095

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

Ser Arg Val Pro Leu Gln Asp Ser Ala Cys Glu Asn Thr Gly Asp
1115 1120 1125

Ile

<210> 233
<211> 1368
<212> PRT
<213> Mus musculus

<220>
<221> MISC_FEATURE
<222> (1)..(1368)
<223> high fidelity Cas9 polypeptide sequence

<400> 233

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
```

```
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430
```

```
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Ala Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
```

```
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860
```

```
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Ala Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065
```

```
Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275
```

```
Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290


Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305


Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320


Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                1330                1335


Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350


Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365


<210>  234
<211>  1368
<212>  PRT
<213>  S. pyogenes


<220>
<221>  MISC_FEATURE
<222>  (1)..(1368)
<223>  PAM-binding SpEQR Cas9 polypeptide sequence

<400>  234

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60
```

```
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Val Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
```

```
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
```

```
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
```

```
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925
```

```
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Glu  Ser Pro Thr Val Ala  Tyr Ser Val
```

```
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Gln  Tyr Arg Ser
    1325                1330                1335
```

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
    1340   1345   1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
    1355   1360   1365

<210> 235
<211> 1368
<212> PRT
<213> S. pyogenes

<220>
<221> MISC_FEATURE
<222> (1)..(1368)
<223> PAM-binding SpVQR Cas9 polypeptide sequence

<400> 235

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

```
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
```

```
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560
```

```
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
```

```
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990
```

```
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080


Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095


Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110


Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125


Lys Lys  Tyr Gly Gly Phe Val  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140


Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155


Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170


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


Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200
```

```
Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                 1210                 1215


Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                 1225                 1230


Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                 1240                 1245


Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                 1255                 1260


His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                 1270                 1275


Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                 1285                 1290


Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                 1300                 1305


Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320


Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Gln  Tyr Arg Ser
    1325                 1330                 1335


Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350


Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365


<210>  236
<211>  1368
<212>  PRT
<213>  S. pyogenes


<220>
<221>  MISC_FEATURE
```

<222> (1)..(1368)
<223> SpVQR Cas9 polypeptide sequence

<400> 236

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60


Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80


Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95


Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110


His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125


His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140


Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160


Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175


Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190


Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
```

```
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
```

```
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
```

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

```
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
```

```
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Val  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Arg
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260
```

```
His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275


Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290


Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                1300                1305


Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                1315                1320


Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Glu  Tyr Arg Ser
    1325                1330                1335


Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                1345                1350


Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                1360                1365


<210>  237
<211>  1359
<212>  PRT
<213>  Streptococcus macacae


<220>
<221>  MISC_FEATURE
<222>  (1)..(1359)
<223>  SpyMacCas9 polypeptide sequence

<400>  237

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
```

```
Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
```

```
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480
```

```
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

```
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705                 710                 715                 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
        755                 760                 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
    770                 775                 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785                 790                 795                 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
                805                 810                 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
            820                 825                 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
        835                 840                 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
    850                 855                 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865                 870                 875                 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
                885                 890                 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
            900                 905                 910
```

```
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
        915                 920                 925


His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
    930                 935                 940


Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945                 950                 955                 960


Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
                965                 970                 975


Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
            980                 985                 990


Gly Thr Ala Leu Ile Lys Lys Tyr  Pro Lys Leu Glu Ser  Glu Phe Val
        995                 1000                1005


Tyr Gly  Asp Tyr Lys Val Tyr  Asp Val Arg Lys Met  Ile Ala Lys
    1010                1015                1020


Ser Glu  Gln Glu Ile Gly Lys  Ala Thr Ala Lys Tyr  Phe Phe Tyr
    1025                1030                1035


Ser Asn  Ile Met Asn Phe Phe  Lys Thr Glu Ile Thr  Leu Ala Asn
    1040                1045                1050


Gly Glu  Ile Arg Lys Arg Pro  Leu Ile Glu Thr Asn  Gly Glu Thr
    1055                1060                1065


Gly Glu  Ile Val Trp Asp Lys  Gly Arg Asp Phe Ala  Thr Val Arg
    1070                1075                1080


Lys Val  Leu Ser Met Pro Gln  Val Asn Ile Val Lys  Lys Thr Glu
    1085                1090                1095


Ile Gln  Thr Val Gly Gln Asn  Gly Gly Leu Phe Asp  Asp Asn Pro
    1100                1105                1110


Lys Ser  Pro Leu Glu Val Thr  Pro Ser Lys Leu Val  Pro Leu Lys
    1115                1120                1125
```

```
Lys Glu  Leu Asn Pro Lys Lys  Tyr Gly Gly Tyr Gln  Lys Pro Thr
    1130                1135                1140

Thr Ala  Tyr Pro Val Leu Leu  Ile Thr Asp Thr Lys  Gln Leu Ile
    1145                1150                1155

Pro Ile  Ser Val Met Asn Lys  Lys Gln Phe Glu Gln  Asn Pro Val
    1160                1165                1170

Lys Phe  Leu Arg Asp Arg Gly  Tyr Gln Gln Val Gly  Lys Asn Asp
    1175                1180                1185

Phe Ile  Lys Leu Pro Lys Tyr  Thr Leu Val Asp Ile  Gly Asp Gly
    1190                1195                1200

Ile Lys  Arg Leu Trp Ala Ser  Ser Lys Glu Ile His  Lys Gly Asn
    1205                1210                1215

Gln Leu  Val Val Ser Lys Lys  Ser Gln Ile Leu Leu  Tyr His Ala
    1220                1225                1230

His His  Leu Asp Ser Asp Leu  Ser Asn Asp Tyr Leu  Gln Asn His
    1235                1240                1245

Asn Gln  Gln Phe Asp Val Leu  Phe Asn Glu Ile Ile  Ser Phe Ser
    1250                1255                1260

Lys Lys  Cys Lys Leu Gly Lys  Glu His Ile Gln Lys  Ile Glu Asn
    1265                1270                1275

Val Tyr  Ser Asn Lys Lys Asn  Ser Ala Ser Ile Glu  Glu Leu Ala
    1280                1285                1290

Glu Ser  Phe Ile Lys Leu Leu  Gly Phe Thr Gln Leu  Gly Ala Thr
    1295                1300                1305

Ser Pro  Phe Asn Phe Leu Gly  Val Lys Leu Asn Gln  Lys Gln Tyr
    1310                1315                1320
```

Lys Gly Lys Lys Asp Tyr Ile Leu Pro Cys Thr Glu Gly Thr Leu
    1325   1330   1335

Ile Arg Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Val Asp Leu
    1340   1345   1350

Ser Lys Ile Gly Glu Asp
    1355

<210> 238
<211> 1410
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: CP5 polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(341)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (342)..(362)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (363)..(1388)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1389)..(1410)
<223> Bipartite nuclear localization sequence

<400> 238

Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met
1   5   10   15

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

Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp
        35   40   45

```
Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln
    50                  55                  60

Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
65                  70                  75                  80

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

Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Met Gln Pro Thr Val
            100                 105                 110

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

Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg
    130                 135                 140

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

Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
                165                 170                 175

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Lys Phe
            180                 185                 190

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe
        195                 200                 205

Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp
    210                 215                 220

Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp
225                 230                 235                 240

Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala
                245                 250                 255

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

```
            260                 265                 270

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

Thr Asn Leu Gly Ala Pro Arg Ala Phe Lys Tyr Phe Asp Thr Thr Ile
    290                 295                 300

Ala Arg Lys Glu Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
305                 310                 315                 320

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser
                325                 330                 335

Gln Leu Gly Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
            340                 345                 350

Ser Gly Gly Ser Gly Gly Ser Gly Gly Met Asp Lys Lys Tyr Ser Ile
        355                 360                 365

Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp
    370                 375                 380

Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp
385                 390                 395                 400

Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser
                405                 410                 415

Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg
            420                 425                 430

Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser
        435                 440                 445

Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu
    450                 455                 460

Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe
465                 470                 475                 480
```

```
Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile
                485                 490                 495

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

Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His
        515                 520                 525

Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys
    530                 535                 540

Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn
545                 550                 555                 560

Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg
                565                 570                 575

Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly
            580                 585                 590

Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly
        595                 600                 605

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

Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu
625                 630                 635                 640

Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn
                645                 650                 655

Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu
            660                 665                 670

Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu
        675                 680                 685

His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu
```

```
    690                 695                 700

Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr
705                 710                 715                 720

Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe
                725                 730                 735

Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val
            740                 745                 750

Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn
        755                 760                 765

Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu
    770                 775                 780

Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys
785                 790                 795                 800

Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu
                805                 810                 815

Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu
            820                 825                 830

Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser
        835                 840                 845

Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro
    850                 855                 860

Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr
865                 870                 875                 880

Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg
                885                 890                 895

Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu
            900                 905                 910
```

```
Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp
        915                 920                 925


Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val
    930                 935                 940


Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys
945                 950                 955                 960


Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile
                965                 970                 975


Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met
            980                 985                 990


Ile Glu Glu Arg Leu Lys Thr Tyr  Ala His Leu Phe Asp  Asp Lys Val
        995                 1000                1005


Met Lys  Gln Leu Lys Arg Arg  Arg Tyr Thr Gly Trp  Gly Arg Leu
    1010                1015                1020


Ser Arg  Lys Leu Ile Asn Gly  Ile Arg Asp Lys Gln  Ser Gly Lys
    1025                1030                1035


Thr Ile  Leu Asp Phe Leu Lys  Ser Asp Gly Phe Ala  Asn Arg Asn
    1040                1045                1050


Phe Met  Gln Leu Ile His Asp  Asp Ser Leu Thr Phe  Lys Glu Asp
    1055                1060                1065


Ile Gln  Lys Ala Gln Val Ser  Gly Gln Gly Asp Ser  Leu His Glu
    1070                1075                1080


His Ile  Ala Asn Leu Ala Gly  Ser Pro Ala Ile Lys  Lys Gly Ile
    1085                1090                1095


Leu Gln  Thr Val Lys Val Val  Asp Glu Leu Val Lys  Val Met Gly
    1100                1105                1110


Arg His  Lys Pro Glu Asn Ile  Val Ile Glu Met Ala  Arg Glu Asn
```

```
    1115                1120                1125

Gln Thr  Thr Gln Lys Gly Gln  Lys Asn Ser Arg Glu  Arg Met Lys
    1130                1135                1140

Arg Ile  Glu Glu Gly Ile Lys  Glu Leu Gly Ser Gln  Ile Leu Lys
    1145                1150                1155

Glu His  Pro Val Glu Asn Thr  Gln Leu Gln Asn Glu  Lys Leu Tyr
    1160                1165                1170

Leu Tyr  Tyr Leu Gln Asn Gly  Arg Asp Met Tyr Val  Asp Gln Glu
    1175                1180                1185

Leu Asp  Ile Asn Arg Leu Ser  Asp Tyr Asp Val Asp  His Ile Val
    1190                1195                1200

Pro Gln  Ser Phe Leu Lys Asp  Asp Ser Ile Asp Asn  Lys Val Leu
    1205                1210                1215

Thr Arg  Ser Asp Lys Asn Arg  Gly Lys Ser Asp Asn  Val Pro Ser
    1220                1225                1230

Glu Glu  Val Val Lys Lys Met  Lys Asn Tyr Trp Arg  Gln Leu Leu
    1235                1240                1245

Asn Ala  Lys Leu Ile Thr Gln  Arg Lys Phe Asp Asn  Leu Thr Lys
    1250                1255                1260

Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp Lys Ala  Gly Phe Ile
    1265                1270                1275

Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile Thr Lys  His Val Ala
    1280                1285                1290

Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp
    1295                1300                1305

Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu
    1310                1315                1320
```

Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
    1325   1330   1335

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
    1340   1345   1350

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu
    1355   1360   1365

Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
    1370   1375   1380

Ala Lys Ser Glu Gln Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly
    1385   1390   1395

Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
    1400   1405   1410

<210> 239
<211> 1302
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Cas12c1 polypeptide sequence

<400> 239

Met Gln Thr Lys Lys Thr His Leu His Leu Ile Ser Ala Lys Ala Ser
1   5   10   15

Arg Lys Tyr Arg Arg Thr Ile Ala Cys Leu Ser Asp Thr Ala Lys Lys
            20   25   30

Asp Leu Glu Arg Arg Lys Gln Ser Gly Ala Ala Asp Pro Ala Gln Glu
        35   40   45

Leu Ser Cys Leu Lys Thr Ile Lys Phe Lys Leu Glu Val Pro Glu Gly
    50   55   60

Ser Lys Leu Pro Ser Phe Asp Arg Ile Ser Gln Ile Tyr Asn Ala Leu
65   70   75   80

```
Glu Thr Ile Glu Lys Gly Ser Leu Ser Tyr Leu Leu Phe Ala Leu Ile
                85                  90                  95

Leu Ser Gly Phe Arg Ile Phe Pro Asn Ser Ser Ala Ala Lys Thr Phe
            100                 105                 110

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

Glu Ile Phe Gly Glu Met Val Lys Asn Phe Ile Pro Ser Glu Leu Glu
    130                 135                 140

Ser Ile Leu Lys Lys Gly Arg Arg Lys Asn Asn Lys Asp Trp Thr Glu
145                 150                 155                 160

Glu Asn Ile Lys Arg Val Leu Asn Ser Glu Phe Gly Arg Lys Asn Ser
                165                 170                 175

Glu Gly Ser Ser Ala Leu Phe Asp Ser Phe Leu Ser Lys Phe Ser Gln
            180                 185                 190

Glu Leu Phe Arg Lys Phe Asp Ser Trp Asn Glu Val Asn Lys Lys Tyr
        195                 200                 205

Leu Glu Ala Ala Glu Leu Leu Asp Ser Met Leu Ala Ser Tyr Gly Pro
    210                 215                 220

Phe Asp Ser Val Cys Lys Met Ile Gly Asp Ser Asp Ser Arg Asn Ser
225                 230                 235                 240

Leu Pro Asp Lys Ser Thr Ile Ala Phe Thr Asn Asn Ala Glu Ile Thr
                245                 250                 255

Val Asp Ile Glu Ser Ser Val Met Pro Tyr Met Ala Ile Ala Ala Leu
            260                 265                 270

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

Val Gln Ser His Leu Thr Thr Thr Asn Gly Asn Gly Leu Ser Trp Phe
```

```
    290                 295                 300

Phe Lys Phe Gly Leu Asp Leu Ile Arg Lys Ala Pro Val Ser Ser Lys
305                 310                 315                 320

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

Pro Asp Asp Lys Leu Asp Gly Leu Lys Phe Ile Lys Glu Ala Cys Glu
            340                 345                 350

Ala Leu Pro Glu Ala Ser Leu Leu Cys Gly Glu Lys Gly Glu Leu Leu
        355                 360                 365

Gly Tyr Gln Asp Phe Arg Thr Ser Phe Ala Gly His Ile Asp Ser Trp
    370                 375                 380

Val Ala Asn Tyr Val Asn Arg Leu Phe Glu Leu Ile Glu Leu Val Asn
385                 390                 395                 400

Gln Leu Pro Glu Ser Ile Lys Leu Pro Ser Ile Leu Thr Gln Lys Asn
                405                 410                 415

His Asn Leu Val Ala Ser Leu Gly Leu Gln Glu Ala Glu Val Ser His
            420                 425                 430

Ser Leu Glu Leu Phe Glu Gly Leu Val Lys Asn Val Arg Gln Thr Leu
        435                 440                 445

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

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

Ile Arg Asn Gln Ile Glu Asn Ala Val Gln Thr Ala Lys Lys Asp Lys
                485                 490                 495

Ile Asp Leu Glu Ser Ala Ile Glu Trp Lys Glu Trp Lys Lys Leu Lys
            500                 505                 510
```

```
Lys Leu Pro Lys Leu Asn Gly Leu Gly Gly Gly Val Pro Lys Gln Gln
        515                 520                 525

Glu Leu Leu Asp Lys Ala Leu Glu Ser Val Lys Gln Ile Arg His Tyr
    530                 535                 540

Gln Arg Ile Asp Phe Glu Arg Val Ile Gln Trp Ala Val Asn Glu His
545                 550                 555                 560

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

Ile Asn Lys Glu Ser Ser Thr Asp Phe Ala Ala Lys Glu Asn Ala Val
            580                 585                 590

Arg Phe Leu Leu Glu Gly Ile Gly Ala Ala Ala Arg Gly Lys Thr Asp
        595                 600                 605

Ser Val Ser Lys Ala Ala Tyr Asn Trp Phe Val Val Asn Asn Phe Leu
    610                 615                 620

Ala Lys Lys Asp Leu Asn Arg Tyr Phe Ile Asn Cys Gln Gly Cys Ile
625                 630                 635                 640

Tyr Lys Pro Pro Tyr Ser Lys Arg Arg Ser Leu Ala Phe Ala Leu Arg
                645                 650                 655

Ser Asp Asn Lys Asp Thr Ile Glu Val Val Trp Glu Lys Phe Glu Thr
            660                 665                 670

Phe Tyr Lys Glu Ile Ser Lys Glu Ile Glu Lys Phe Asn Ile Phe Ser
        675                 680                 685

Gln Glu Phe Gln Thr Phe Leu His Leu Glu Asn Leu Arg Met Lys Leu
    690                 695                 700

Leu Leu Arg Arg Ile Gln Lys Pro Ile Pro Ala Glu Ile Ala Phe Phe
705                 710                 715                 720

Ser Leu Pro Gln Glu Tyr Tyr Asp Ser Leu Pro Pro Asn Val Ala Phe
```

```
                725                 730                 735

Leu Ala Leu Asn Gln Glu Ile Thr Pro Ser Glu Tyr Ile Thr Gln Phe
            740                 745                 750

Asn Leu Tyr Ser Ser Phe Leu Asn Gly Asn Leu Ile Leu Leu Arg Arg
        755                 760                 765

Ser Arg Ser Tyr Leu Arg Ala Lys Phe Ser Trp Val Gly Asn Ser Lys
    770                 775                 780

Leu Ile Tyr Ala Ala Lys Glu Ala Arg Leu Trp Lys Ile Pro Asn Ala
785                 790                 795                 800

Tyr Trp Lys Ser Asp Glu Trp Lys Met Ile Leu Asp Ser Asn Val Leu
                805                 810                 815

Val Phe Asp Lys Ala Gly Asn Val Leu Pro Ala Pro Thr Leu Lys Lys
            820                 825                 830

Val Cys Glu Arg Glu Gly Asp Leu Arg Leu Phe Tyr Pro Leu Leu Arg
        835                 840                 845

Gln Leu Pro His Asp Trp Cys Tyr Arg Asn Pro Phe Val Lys Ser Val
    850                 855                 860

Gly Arg Glu Lys Asn Val Ile Glu Val Asn Lys Glu Gly Glu Pro Lys
865                 870                 875                 880

Val Ala Ser Ala Leu Pro Gly Ser Leu Phe Arg Leu Ile Gly Pro Ala
                885                 890                 895

Pro Phe Lys Ser Leu Leu Asp Asp Cys Phe Phe Asn Pro Leu Asp Lys
            900                 905                 910

Asp Leu Arg Glu Cys Met Leu Ile Val Asp Gln Glu Ile Ser Gln Lys
        915                 920                 925

Val Glu Ala Gln Lys Val Glu Ala Ser Leu Glu Ser Cys Thr Tyr Ser
    930                 935                 940
```

```
Ile Ala Val Pro Ile Arg Tyr His Leu Glu Glu Pro Lys Val Ser Asn
945                 950                 955                 960


Gln Phe Glu Asn Val Leu Ala Ile Asp Gln Gly Glu Ala Gly Leu Ala
                965                 970                 975


Tyr Ala Val Phe Ser Leu Lys Ser Ile Gly Glu Ala Glu Thr Lys Pro
            980                 985                 990


Ile Ala Val Gly Thr Ile Arg Ile  Pro Ser Ile Arg Arg  Leu Ile His
        995                 1000                1005


Ser Val  Ser Thr Tyr Arg Lys  Lys Lys Gln Arg Leu  Gln Asn Phe
    1010                1015                1020


Lys Gln  Asn Tyr Asp Ser Thr  Ala Phe Ile Met Arg  Glu Asn Val
    1025                1030                1035


Thr Gly  Asp Val Cys Ala Lys  Ile Val Gly Leu Met  Lys Glu Phe
    1040                1045                1050


Asn Ala  Phe Pro Val Leu Glu  Tyr Asp Val Lys Asn  Leu Glu Ser
    1055                1060                1065


Gly Ser  Arg Gln Leu Ser Ala  Val Tyr Lys Ala Val  Asn Ser His
    1070                1075                1080


Phe Leu  Tyr Phe Lys Glu Pro  Gly Arg Asp Ala Leu  Arg Lys Gln
    1085                1090                1095


Leu Trp  Tyr Gly Gly Asp Ser  Trp Thr Ile Asp Gly  Ile Glu Ile
    1100                1105                1110


Val Thr  Arg Glu Arg Lys Glu  Asp Gly Lys Glu Gly  Val Glu Lys
    1115                1120                1125


Ile Val  Pro Leu Lys Val Phe  Pro Gly Arg Ser Val  Ser Ala Arg
    1130                1135                1140


Phe Thr  Ser Lys Thr Cys Ser  Cys Cys Gly Arg Asn  Val Phe Asp
```

```
    1145                1150                1155

Trp Leu  Phe Thr Glu Lys Lys  Ala Lys Thr Asn Lys  Lys Phe Asn
    1160                1165                1170

Val Asn  Ser Lys Gly Glu Leu  Thr Thr Ala Asp Gly  Val Ile Gln
    1175                1180                1185

Leu Phe  Glu Ala Asp Arg Ser  Lys Gly Pro Lys Phe  Tyr Ala Arg
    1190                1195                1200

Arg Lys  Glu Arg Thr Pro Leu  Thr Lys Pro Ile Ala  Lys Gly Ser
    1205                1210                1215

Tyr Ser  Leu Glu Glu Ile Glu  Arg Arg Val Arg Thr  Asn Leu Arg
    1220                1225                1230

Arg Ala  Pro Lys Ser Lys Gln  Ser Arg Asp Thr Ser  Gln Ser Gln
    1235                1240                1245

Tyr Phe  Cys Val Tyr Lys Asp  Cys Ala Leu His Phe  Ser Gly Met
    1250                1255                1260

Gln Ala  Asp Glu Asn Ala Ala  Ile Asn Ile Gly Arg  Arg Phe Leu
    1265                1270                1275

Thr Ala  Leu Arg Lys Asn Arg  Arg Ser Asp Phe Pro  Ser Asn Val
    1280                1285                1290

Lys Ile  Ser Asp Arg Leu Leu  Asp Asn
    1295                1300


<210>  240
<211>  1218
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: Cas12c2 polypeptide sequence

<400>  240

Met Thr Lys His Ser Ile Pro Leu His Ala Phe Arg Asn Ser Gly Ala
```

```
1               5                   10                  15

Asp Ala Arg Lys Trp Lys Gly Arg Ile Ala Leu Leu Ala Lys Arg Gly
            20                  25                  30

Lys Glu Thr Met Arg Thr Leu Gln Phe Pro Leu Glu Met Ser Glu Pro
        35                  40                  45

Glu Ala Ala Ala Ile Asn Thr Thr Pro Phe Ala Val Ala Tyr Asn Ala
    50                  55                  60

Ile Glu Gly Thr Gly Lys Gly Thr Leu Phe Asp Tyr Trp Ala Lys Leu
65                  70                  75                  80

His Leu Ala Gly Phe Arg Phe Phe Pro Ser Gly Gly Ala Ala Thr Ile
                85                  90                  95

Phe Arg Gln Gln Ala Val Phe Glu Asp Ala Ser Trp Asn Ala Ala Phe
            100                 105                 110

Cys Gln Gln Ser Gly Lys Asp Trp Pro Trp Leu Val Pro Ser Lys Leu
        115                 120                 125

Tyr Glu Arg Phe Thr Lys Ala Pro Arg Glu Val Ala Lys Lys Asp Gly
    130                 135                 140

Ser Lys Lys Ser Ile Glu Phe Thr Gln Glu Asn Val Ala Asn Glu Ser
145                 150                 155                 160

His Val Ser Leu Val Gly Ala Ser Ile Thr Asp Lys Thr Pro Glu Asp
                165                 170                 175

Gln Lys Glu Phe Phe Leu Lys Met Ala Gly Ala Leu Ala Glu Lys Phe
            180                 185                 190

Asp Ser Trp Lys Ser Ala Asn Glu Asp Arg Ile Val Ala Met Lys Val
        195                 200                 205

Ile Asp Glu Phe Leu Lys Ser Glu Gly Leu His Leu Pro Ser Leu Glu
    210                 215                 220
```

```
Asn Ile Ala Val Lys Cys Ser Val Glu Thr Lys Pro Asp Asn Ala Thr
225                 230                 235                 240

Val Ala Trp His Asp Ala Pro Met Ser Gly Val Gln Asn Leu Ala Ile
                245                 250                 255

Gly Val Phe Ala Thr Cys Ala Ser Arg Ile Asp Asn Ile Tyr Asp Leu
            260                 265                 270

Asn Gly Gly Lys Leu Ser Lys Leu Ile Gln Glu Ser Ala Thr Thr Pro
        275                 280                 285

Asn Val Thr Ala Leu Ser Trp Leu Phe Gly Lys Gly Leu Glu Tyr Phe
    290                 295                 300

Arg Thr Thr Asp Ile Asp Thr Ile Met Gln Asp Phe Asn Ile Pro Ala
305                 310                 315                 320

Ser Ala Lys Glu Ser Ile Lys Pro Leu Val Glu Ser Ala Gln Ala Ile
                325                 330                 335

Pro Thr Met Thr Val Leu Gly Lys Lys Asn Tyr Ala Pro Phe Arg Pro
            340                 345                 350

Asn Phe Gly Gly Lys Ile Asp Ser Trp Ile Ala Asn Tyr Ala Ser Arg
        355                 360                 365

Leu Met Leu Leu Asn Asp Ile Leu Glu Gln Ile Glu Pro Gly Phe Glu
    370                 375                 380

Leu Pro Gln Ala Leu Leu Asp Asn Glu Thr Leu Met Ser Gly Ile Asp
385                 390                 395                 400

Met Thr Gly Asp Glu Leu Lys Glu Leu Ile Glu Ala Val Tyr Ala Trp
                405                 410                 415

Val Asp Ala Ala Lys Gln Gly Leu Ala Thr Leu Leu Gly Arg Gly Gly
            420                 425                 430

Asn Val Asp Asp Ala Val Gln Thr Phe Glu Gln Phe Ser Ala Met Met
```

```
        435                 440                 445

Asp Thr Leu Asn Gly Thr Leu Asn Thr Ile Ser Ala Arg Tyr Val Arg
    450                 455                 460

Ala Val Glu Met Ala Gly Lys Asp Glu Ala Arg Leu Glu Lys Leu Ile
465                 470                 475                 480

Glu Cys Lys Phe Asp Ile Pro Lys Trp Cys Lys Ser Val Pro Lys Leu
                485                 490                 495

Val Gly Ile Ser Gly Gly Leu Pro Lys Val Glu Glu Glu Ile Lys Val
            500                 505                 510

Met Asn Ala Ala Phe Lys Asp Val Arg Ala Arg Met Phe Val Arg Phe
        515                 520                 525

Glu Glu Ile Ala Ala Tyr Val Ala Ser Lys Gly Ala Gly Met Asp Val
    530                 535                 540

Tyr Asp Ala Leu Glu Lys Arg Glu Leu Glu Gln Ile Lys Lys Leu Lys
545                 550                 555                 560

Ser Ala Val Pro Glu Arg Ala His Ile Gln Ala Tyr Arg Ala Val Leu
                565                 570                 575

His Arg Ile Gly Arg Ala Val Gln Asn Cys Ser Glu Lys Thr Lys Gln
            580                 585                 590

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

His Leu Asn Asn Phe Ile Phe Asn Gln Lys Gly Ala Ile Tyr Arg Ser
    610                 615                 620

Pro Phe Asp Arg Ser Arg His Ala Pro Tyr Gln Leu His Ala Asp Lys
625                 630                 635                 640

Leu Leu Lys Asn Asp Trp Leu Glu Leu Leu Ala Glu Ile Ser Ala Thr
                645                 650                 655
```

```
Leu Met Ala Ser Glu Ser Thr Glu Gln Met Glu Asp Ala Leu Arg Leu
            660                 665                 670

Glu Arg Thr Arg Leu Gln Leu Gln Leu Ser Gly Leu Pro Asp Trp Glu
        675                 680                 685

Tyr Pro Ala Ser Leu Ala Lys Pro Asp Ile Glu Val Glu Ile Gln Thr
    690                 695                 700

Ala Leu Lys Met Gln Leu Ala Lys Asp Thr Val Thr Ser Asp Val Leu
705                 710                 715                 720

Gln Arg Ala Phe Asn Leu Tyr Ser Ser Val Leu Ser Gly Leu Thr Phe
                725                 730                 735

Lys Leu Leu Arg Arg Ser Phe Ser Leu Lys Met Arg Phe Ser Val Ala
            740                 745                 750

Asp Thr Thr Gln Leu Ile Tyr Val Pro Lys Val Cys Asp Trp Ala Ile
        755                 760                 765

Pro Lys Gln Tyr Leu Gln Ala Glu Gly Glu Ile Gly Ile Ala Ala Arg
    770                 775                 780

Val Val Thr Glu Ser Ser Pro Ala Lys Met Val Thr Glu Val Glu Met
785                 790                 795                 800

Lys Glu Pro Lys Ala Leu Gly His Phe Met Gln Gln Ala Pro His Asp
                805                 810                 815

Trp Tyr Phe Asp Ala Ser Leu Gly Gly Thr Gln Val Ala Gly Arg Ile
            820                 825                 830

Val Glu Lys Gly Lys Glu Val Gly Lys Glu Arg Lys Leu Val Gly Tyr
        835                 840                 845

Arg Met Arg Gly Asn Ser Ala Tyr Lys Thr Val Leu Asp Lys Ser Leu
    850                 855                 860

Val Gly Asn Thr Glu Leu Ser Gln Cys Ser Met Ile Ile Glu Ile Pro
```

```
865                 870                 875                 880

Tyr Thr Gln Thr Val Asp Ala Asp Phe Arg Ala Gln Val Gln Ala Gly
                885                 890                 895

Leu Pro Lys Val Ser Ile Asn Leu Pro Val Lys Glu Thr Ile Thr Ala
            900                 905                 910

Ser Asn Lys Asp Glu Gln Met Leu Phe Asp Arg Phe Val Ala Ile Asp
        915                 920                 925

Leu Gly Glu Arg Gly Leu Gly Tyr Ala Val Phe Asp Ala Lys Thr Leu
    930                 935                 940

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

Leu Asn Arg Thr His His Tyr Glu Gln Arg Pro Asn Gln Arg Gln Lys
                965                 970                 975

Phe Gln Ala Lys Phe Asn Val Asn Leu Ser Glu Leu Arg Glu Asn Thr
            980                 985                 990

Val Gly Asp Val Cys His Gln Ile  Asn Arg Ile Cys Ala  Tyr Tyr Asn
        995                 1000                1005

Ala Phe  Pro Val Leu Glu Tyr  Met Val Pro Asp Arg  Leu Asp Lys
    1010                1015                1020

Gln Leu  Lys Ser Val Tyr Glu  Ser Val Thr Asn Arg  Tyr Ile Trp
    1025                1030                1035

Ser Ser  Thr Asp Ala His Lys  Ser Ala Arg Val Gln  Phe Trp Leu
    1040                1045                1050

Gly Gly  Glu Thr Trp Glu His  Pro Tyr Leu Lys Ser  Ala Lys Asp
    1055                1060                1065

Lys Lys  Pro Leu Val Leu Ser  Pro Gly Arg Gly Ala  Ser Gly Lys
    1070                1075                1080
```

```
Gly Thr  Ser Gln Thr Cys Ser  Cys Cys Gly Arg Asn  Pro Phe Asp
    1085                1090                1095


Leu Ile  Lys Asp Met Lys Pro  Arg Ala Lys Ile Ala  Val Val Asp
    1100                1105                1110


Gly Lys  Ala Lys Leu Glu Asn  Ser Glu Leu Lys Leu  Phe Glu Arg
    1115                1120                1125


Asn Leu  Glu Ser Lys Asp Asp  Met Leu Ala Arg Arg  His Arg Asn
    1130                1135                1140


Glu Arg  Ala Gly Met Glu Gln  Pro Leu Thr Pro Gly  Asn Tyr Thr
    1145                1150                1155


Val Asp  Glu Ile Lys Ala Leu  Leu Arg Ala Asn Leu  Arg Arg Ala
    1160                1165                1170


Pro Lys  Asn Arg Arg Thr Lys  Asp Thr Thr Val Ser  Glu Tyr His
    1175                1180                1185


Cys Val  Phe Ser Asp Cys Gly  Lys Thr Met His Ala  Asp Glu Asn
    1190                1195                1200


Ala Ala  Val Asn Ile Gly Gly  Lys Phe Ile Ala Asp  Ile Glu Lys
    1205                1210                1215


<210>  241
<211>  1252
<212>  PRT
<213>  Oleiphilus sp. HI0009


<220>
<221>  MISC_FEATURE
<222>  (1)..(1252)
<223>  OspCas12c polypeptide sequence

<400>  241

Met Thr Lys Leu Arg His Arg Gln Lys Lys Leu Thr His Asp Trp Ala
1               5                  10                  15
```

```
Gly Ser Lys Lys Arg Glu Val Leu Gly Ser Asn Gly Lys Leu Gln Asn
            20                  25                  30

Pro Leu Leu Met Pro Val Lys Lys Gly Gln Val Thr Glu Phe Arg Lys
        35                  40                  45

Ala Phe Ser Ala Tyr Ala Arg Ala Thr Lys Gly Glu Met Thr Asp Gly
    50                  55                  60

Arg Lys Asn Met Phe Thr His Ser Phe Glu Pro Phe Lys Thr Lys Pro
65                  70                  75                  80

Ser Leu His Gln Cys Glu Leu Ala Asp Lys Ala Tyr Gln Ser Leu His
                85                  90                  95

Ser Tyr Leu Pro Gly Ser Leu Ala His Phe Leu Leu Ser Ala His Ala
            100                 105                 110

Leu Gly Phe Arg Ile Phe Ser Lys Ser Gly Glu Ala Thr Ala Phe Gln
        115                 120                 125

Ala Ser Ser Lys Ile Glu Ala Tyr Glu Ser Lys Leu Ala Ser Glu Leu
    130                 135                 140

Ala Cys Val Asp Leu Ser Ile Gln Asn Leu Thr Ile Ser Thr Leu Phe
145                 150                 155                 160

Asn Ala Leu Thr Thr Ser Val Arg Gly Lys Gly Glu Glu Thr Ser Ala
                165                 170                 175

Asp Pro Leu Ile Ala Arg Phe Tyr Thr Leu Leu Thr Gly Lys Pro Leu
            180                 185                 190

Ser Arg Asp Thr Gln Gly Pro Glu Arg Asp Leu Ala Glu Val Ile Ser
        195                 200                 205

Arg Lys Ile Ala Ser Ser Phe Gly Thr Trp Lys Glu Met Thr Ala Asn
    210                 215                 220

Pro Leu Gln Ser Leu Gln Phe Phe Glu Glu Glu Leu His Ala Leu Asp
225                 230                 235                 240
```

```
Ala Asn Val Ser Leu Ser Pro Ala Phe Asp Val Leu Ile Lys Met Asn
                245                 250                 255

Asp Leu Gln Gly Asp Leu Lys Asn Arg Thr Ile Val Phe Asp Pro Asp
            260                 265                 270

Ala Pro Val Phe Glu Tyr Asn Ala Glu Asp Pro Ala Asp Ile Ile Ile
        275                 280                 285

Lys Leu Thr Ala Arg Tyr Ala Lys Glu Ala Val Ile Lys Asn Gln Asn
    290                 295                 300

Val Gly Asn Tyr Val Lys Asn Ala Ile Thr Thr Thr Asn Ala Asn Gly
305                 310                 315                 320

Leu Gly Trp Leu Leu Asn Lys Gly Leu Ser Leu Leu Pro Val Ser Thr
                325                 330                 335

Asp Asp Glu Leu Leu Glu Phe Ile Gly Val Glu Arg Ser His Pro Ser
            340                 345                 350

Cys His Ala Leu Ile Glu Leu Ile Ala Gln Leu Glu Ala Pro Glu Leu
        355                 360                 365

Phe Glu Lys Asn Val Phe Ser Asp Thr Arg Ser Glu Val Gln Gly Met
    370                 375                 380

Ile Asp Ser Ala Val Ser Asn His Ile Ala Arg Leu Ser Ser Ser Arg
385                 390                 395                 400

Asn Ser Leu Ser Met Asp Ser Glu Glu Leu Glu Arg Leu Ile Lys Ser
                405                 410                 415

Phe Gln Ile His Thr Pro His Cys Ser Leu Phe Ile Gly Ala Gln Ser
            420                 425                 430

Leu Ser Gln Gln Leu Glu Ser Leu Pro Glu Ala Leu Gln Ser Gly Val
        435                 440                 445
```

```
Asn Ser Ala Asp Ile Leu Leu Gly Ser Thr Gln Tyr Met Leu Thr Asn
    450                 455                 460

Ser Leu Val Glu Glu Ser Ile Ala Thr Tyr Gln Arg Thr Leu Asn Arg
465                 470                 475                 480

Ile Asn Tyr Leu Ser Gly Val Ala Gly Gln Ile Asn Gly Ala Ile Lys
                485                 490                 495

Arg Lys Ala Ile Asp Gly Glu Lys Ile His Leu Pro Ala Ala Trp Ser
            500                 505                 510

Glu Leu Ile Ser Leu Pro Phe Ile Gly Gln Pro Val Ile Asp Val Glu
        515                 520                 525

Ser Asp Leu Ala His Leu Lys Asn Gln Tyr Gln Thr Leu Ser Asn Glu
    530                 535                 540

Phe Asp Thr Leu Ile Ser Ala Leu Gln Lys Asn Phe Asp Leu Asn Phe
545                 550                 555                 560

Asn Lys Ala Leu Leu Asn Arg Thr Gln His Phe Glu Ala Met Cys Arg
                565                 570                 575

Ser Thr Lys Lys Asn Ala Leu Ser Lys Pro Glu Ile Val Ser Tyr Arg
            580                 585                 590

Asp Leu Leu Ala Arg Leu Thr Ser Cys Leu Tyr Arg Gly Ser Leu Val
        595                 600                 605

Leu Arg Arg Ala Gly Ile Glu Val Leu Lys Lys His Lys Ile Phe Glu
    610                 615                 620

Ser Asn Ser Glu Leu Arg Glu His Val His Glu Arg Lys His Phe Val
625                 630                 635                 640

Phe Val Ser Pro Leu Asp Arg Lys Ala Lys Lys Leu Leu Arg Leu Thr
                645                 650                 655

Asp Ser Arg Pro Asp Leu Leu His Val Ile Asp Glu Ile Leu Gln His
            660                 665                 670
```

```
Asp Asn Leu Glu Asn Lys Asp Arg Glu Ser Leu Trp Leu Val Arg Ser
        675                 680                 685

Gly Tyr Leu Leu Ala Gly Leu Pro Asp Gln Leu Ser Ser Ser Phe Ile
    690                 695                 700

Asn Leu Pro Ile Ile Thr Gln Lys Gly Asp Arg Arg Leu Ile Asp Leu
705                 710                 715                 720

Ile Gln Tyr Asp Gln Ile Asn Arg Asp Ala Phe Val Met Leu Val Thr
                725                 730                 735

Ser Ala Phe Lys Ser Asn Leu Ser Gly Leu Gln Tyr Arg Ala Asn Lys
            740                 745                 750

Gln Ser Phe Val Val Thr Arg Thr Leu Ser Pro Tyr Leu Gly Ser Lys
        755                 760                 765

Leu Val Tyr Val Pro Lys Asp Lys Asp Trp Leu Val Pro Ser Gln Met
    770                 775                 780

Phe Glu Gly Arg Phe Ala Asp Ile Leu Gln Ser Asp Tyr Met Val Trp
785                 790                 795                 800

Lys Asp Ala Gly Arg Leu Cys Val Ile Asp Thr Ala Lys His Leu Ser
                805                 810                 815

Asn Ile Lys Lys Ser Val Phe Ser Ser Glu Glu Val Leu Ala Phe Leu
            820                 825                 830

Arg Glu Leu Pro His Arg Thr Phe Ile Gln Thr Glu Val Arg Gly Leu
        835                 840                 845

Gly Val Asn Val Asp Gly Ile Ala Phe Asn Asn Gly Asp Ile Pro Ser
    850                 855                 860

Leu Lys Thr Phe Ser Asn Cys Val Gln Val Lys Val Ser Arg Thr Asn
865                 870                 875                 880
```

```
Thr Ser Leu Val Gln Thr Leu Asn Arg Trp Phe Glu Gly Gly Lys Val
                885                 890                 895

Ser Pro Pro Ser Ile Gln Phe Glu Arg Ala Tyr Tyr Lys Lys Asp Asp
            900                 905                 910

Gln Ile His Glu Asp Ala Ala Lys Arg Lys Ile Arg Phe Gln Met Pro
        915                 920                 925

Ala Thr Glu Leu Val His Ala Ser Asp Asp Ala Gly Trp Thr Pro Ser
    930                 935                 940

Tyr Leu Leu Gly Ile Asp Pro Gly Glu Tyr Gly Met Gly Leu Ser Leu
945                 950                 955                 960

Val Ser Ile Asn Asn Gly Glu Val Leu Asp Ser Gly Phe Ile His Ile
                965                 970                 975

Asn Ser Leu Ile Asn Phe Ala Ser Lys Lys Ser Asn His Gln Thr Lys
            980                 985                 990

Val Val Pro Arg Gln Gln Tyr Lys  Ser Pro Tyr Ala Asn  Tyr Leu Glu
        995                 1000                1005

Gln Ser  Lys Asp Ser Ala Ala  Gly Asp Ile Ala His  Ile Leu Asp
    1010                1015                1020

Arg Leu  Ile Tyr Lys Leu Asn  Ala Leu Pro Val Phe  Glu Ala Leu
    1025                1030                1035

Ser Gly  Asn Ser Gln Ser Ala  Ala Asp Gln Val Trp  Thr Lys Val
    1040                1045                1050

Leu Ser  Phe Tyr Thr Trp Gly  Asp Asn Asp Ala Gln  Asn Ser Ile
    1055                1060                1065

Arg Lys  Gln His Trp Phe Gly  Ala Ser His Trp Asp  Ile Lys Gly
    1070                1075                1080

Met Leu  Arg Gln Pro Pro Thr  Glu Lys Lys Pro Lys  Pro Tyr Ile
    1085                1090                1095
```

```
Ala Phe  Pro Gly Ser Gln Val  Ser Ser Tyr Gly Asn  Ser Gln Arg
    1100                1105                1110

Cys Ser  Cys Cys Gly Arg Asn  Pro Ile Glu Gln Leu  Arg Glu Met
    1115                1120                1125

Ala Lys  Asp Thr Ser Ile Lys  Glu Leu Lys Ile Arg  Asn Ser Glu
    1130                1135                1140

Ile Gln  Leu Phe Asp Gly Thr  Ile Lys Leu Phe Asn  Pro Asp Pro
    1145                1150                1155

Ser Thr  Val Ile Glu Arg Arg  Arg His Asn Leu Gly  Pro Ser Arg
    1160                1165                1170

Ile Pro  Val Ala Asp Arg Thr  Phe Lys Asn Ile Ser  Pro Ser Ser
    1175                1180                1185

Leu Glu  Phe Lys Glu Leu Ile  Thr Ile Val Ser Arg  Ser Ile Arg
    1190                1195                1200

His Ser  Pro Glu Phe Ile Ala  Lys Lys Arg Gly Ile  Gly Ser Glu
    1205                1210                1215

Tyr Phe  Cys Ala Tyr Ser Asp  Cys Asn Ser Ser Leu  Asn Ser Glu
    1220                1225                1230

Ala Asn  Ala Ala Ala Asn Val  Ala Gln Lys Phe Gln  Lys Gln Leu
    1235                1240                1245

Phe Phe  Glu Leu
    1250
```

<210> 242
<211> 767
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Cas12g1 polypeptide sequence

<400> 242

```
Met Ala Gln Ala Ser Ser Thr Pro Ala Val Ser Pro Arg Pro Arg Pro
1               5                  10                  15

Arg Tyr Arg Glu Glu Arg Thr Leu Val Arg Lys Leu Leu Pro Arg Pro
            20                  25                  30

Gly Gln Ser Lys Gln Glu Phe Arg Glu Asn Val Lys Lys Leu Arg Lys
        35                  40                  45

Ala Phe Leu Gln Phe Asn Ala Asp Val Ser Gly Val Cys Gln Trp Ala
    50                  55                  60

Ile Gln Phe Arg Pro Arg Tyr Gly Lys Pro Ala Glu Pro Thr Glu Thr
65                  70                  75                  80

Phe Trp Lys Phe Phe Leu Glu Pro Glu Thr Ser Leu Pro Pro Asn Asp
                85                  90                  95

Ser Arg Ser Pro Glu Phe Arg Arg Leu Gln Ala Phe Glu Ala Ala Ala
            100                 105                 110

Gly Ile Asn Gly Ala Ala Ala Leu Asp Asp Pro Ala Phe Thr Asn Glu
        115                 120                 125

Leu Arg Asp Ser Ile Leu Ala Val Ala Ser Arg Pro Lys Thr Lys Glu
    130                 135                 140

Ala Gln Arg Leu Phe Ser Arg Leu Lys Asp Tyr Gln Pro Ala His Arg
145                 150                 155                 160

Met Ile Leu Ala Lys Val Ala Ala Glu Trp Ile Glu Ser Arg Tyr Arg
                165                 170                 175

Arg Ala His Gln Asn Trp Glu Arg Asn Tyr Glu Glu Trp Lys Lys Glu
            180                 185                 190

Lys Gln Glu Trp Glu Gln Asn His Pro Glu Leu Thr Pro Glu Ile Arg
        195                 200                 205
```

```
Glu Ala Phe Asn Gln Ile Phe Gln Gln Leu Glu Val Lys Glu Lys Arg
    210                 215                 220

Val Arg Ile Cys Pro Ala Ala Arg Leu Leu Gln Asn Lys Asp Asn Cys
225                 230                 235                 240

Gln Tyr Ala Gly Lys Asn Lys His Ser Val Leu Cys Asn Gln Phe Asn
                245                 250                 255

Glu Phe Lys Lys Asn His Leu Gln Gly Lys Ala Ile Lys Phe Phe Tyr
            260                 265                 270

Lys Asp Ala Glu Lys Tyr Leu Arg Cys Gly Leu Gln Ser Leu Lys Pro
        275                 280                 285

Asn Val Gln Gly Pro Phe Arg Glu Asp Trp Asn Lys Tyr Leu Arg Tyr
    290                 295                 300

Met Asn Leu Lys Glu Glu Thr Leu Arg Gly Lys Asn Gly Gly Arg Leu
305                 310                 315                 320

Pro His Cys Lys Asn Leu Gly Gln Glu Cys Glu Phe Asn Pro His Thr
                325                 330                 335

Ala Leu Cys Lys Gln Tyr Gln Gln Gln Leu Ser Ser Arg Pro Asp Leu
            340                 345                 350

Val Gln His Asp Glu Leu Tyr Arg Lys Trp Arg Arg Glu Tyr Trp Arg
        355                 360                 365

Glu Pro Arg Lys Pro Val Phe Arg Tyr Pro Ser Val Lys Arg His Ser
    370                 375                 380

Ile Ala Lys Ile Phe Gly Glu Asn Tyr Phe Gln Ala Asp Phe Lys Asn
385                 390                 395                 400

Ser Val Val Gly Leu Arg Leu Asp Ser Met Pro Ala Gly Gln Tyr Leu
                405                 410                 415

Glu Phe Ala Phe Ala Pro Trp Pro Arg Asn Tyr Arg Pro Gln Pro Gly
            420                 425                 430
```

```
Glu Thr Glu Ile Ser Ser Val His Leu His Phe Val Gly Thr Arg Pro
        435                 440                 445

Arg Ile Gly Phe Arg Phe Arg Val Pro His Lys Arg Ser Arg Phe Asp
    450                 455                 460

Cys Thr Gln Glu Glu Leu Asp Glu Leu Arg Ser Arg Thr Phe Pro Arg
465                 470                 475                 480

Lys Ala Gln Asp Gln Lys Phe Leu Glu Ala Ala Arg Lys Arg Leu Leu
                485                 490                 495

Glu Thr Phe Pro Gly Asn Ala Glu Gln Glu Leu Arg Leu Leu Ala Val
            500                 505                 510

Asp Leu Gly Thr Asp Ser Ala Arg Ala Ala Phe Phe Ile Gly Lys Thr
        515                 520                 525

Phe Gln Gln Ala Phe Pro Leu Lys Ile Val Lys Ile Glu Lys Leu Tyr
    530                 535                 540

Glu Gln Trp Pro Asn Gln Lys Gln Ala Gly Asp Arg Arg Asp Ala Ser
545                 550                 555                 560

Ser Lys Gln Pro Arg Pro Gly Leu Ser Arg Asp His Val Gly Arg His
                565                 570                 575

Leu Gln Lys Met Arg Ala Gln Ala Ser Glu Ile Ala Gln Lys Arg Gln
            580                 585                 590

Glu Leu Thr Gly Thr Pro Ala Pro Glu Thr Thr Thr Asp Gln Ala Ala
        595                 600                 605

Lys Lys Ala Thr Leu Gln Pro Phe Asp Leu Arg Gly Leu Thr Val His
    610                 615                 620

Thr Ala Arg Met Ile Arg Asp Trp Ala Arg Leu Asn Ala Arg Gln Ile
625                 630                 635                 640
```

```
Ile Gln Leu Ala Glu Glu Asn Gln Val Asp Leu Ile Val Leu Glu Ser
                645                 650                 655


Leu Arg Gly Phe Arg Pro Pro Gly Tyr Glu Asn Leu Asp Gln Glu Lys
            660                 665                 670


Lys Arg Arg Val Ala Phe Phe Ala His Gly Arg Ile Arg Arg Lys Val
        675                 680                 685


Thr Glu Lys Ala Val Glu Arg Gly Met Arg Val Val Thr Val Pro Tyr
    690                 695                 700


Leu Ala Ser Ser Lys Val Cys Ala Glu Cys Arg Lys Lys Gln Lys Asp
705                 710                 715                 720


Asn Lys Gln Trp Glu Lys Asn Lys Lys Arg Gly Leu Phe Lys Cys Glu
                725                 730                 735


Gly Cys Gly Ser Gln Ala Gln Val Asp Glu Asn Ala Ala Arg Val Leu
            740                 745                 750


Gly Arg Val Phe Trp Gly Glu Ile Glu Leu Pro Thr Ala Ile Pro
        755                 760                 765


<210>  243
<211>  870
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: Cas12h1 polypeptide sequence

<400>  243

Met Lys Val His Glu Ile Pro Arg Ser Gln Leu Leu Lys Ile Lys Gln
1               5                   10                  15


Tyr Glu Gly Ser Phe Val Glu Trp Tyr Arg Asp Leu Gln Glu Asp Arg
            20                  25                  30


Lys Lys Phe Ala Ser Leu Leu Phe Arg Trp Ala Ala Phe Gly Tyr Ala
        35                  40                  45
```

```
Ala Arg Glu Asp Asp Gly Ala Thr Tyr Ile Ser Pro Ser Gln Ala Leu
    50                  55                  60

Leu Glu Arg Arg Leu Leu Leu Gly Asp Ala Glu Asp Val Ala Ile Lys
65                  70                  75                  80

Phe Leu Asp Val Leu Phe Lys Gly Gly Ala Pro Ser Ser Ser Cys Tyr
                85                  90                  95

Ser Leu Phe Tyr Glu Asp Phe Ala Leu Arg Asp Lys Ala Lys Tyr Ser
            100                 105                 110

Gly Ala Lys Arg Glu Phe Ile Glu Gly Leu Ala Thr Met Pro Leu Asp
        115                 120                 125

Lys Ile Ile Glu Arg Ile Arg Gln Asp Glu Gln Leu Ser Lys Ile Pro
    130                 135                 140

Ala Glu Glu Trp Leu Ile Leu Gly Ala Glu Tyr Ser Pro Glu Glu Ile
145                 150                 155                 160

Trp Glu Gln Val Ala Pro Arg Ile Val Asn Val Asp Arg Ser Leu Gly
                165                 170                 175

Lys Gln Leu Arg Glu Arg Leu Gly Ile Lys Cys Arg Arg Pro His Asp
            180                 185                 190

Ala Gly Tyr Cys Lys Ile Leu Met Glu Val Val Ala Arg Gln Leu Arg
        195                 200                 205

Ser His Asn Glu Thr Tyr His Glu Tyr Leu Asn Gln Thr His Glu Met
    210                 215                 220

Lys Thr Lys Val Ala Asn Asn Leu Thr Asn Glu Phe Asp Leu Val Cys
225                 230                 235                 240

Glu Phe Ala Glu Val Leu Glu Glu Lys Asn Tyr Gly Leu Gly Trp Tyr
                245                 250                 255

Val Leu Trp Gln Gly Val Lys Gln Ala Leu Lys Glu Gln Lys Lys Pro
            260                 265                 270
```

```
Thr Lys Ile Gln Ile Ala Val Asp Gln Leu Arg Gln Pro Lys Phe Ala
        275                 280                 285

Gly Leu Leu Thr Ala Lys Trp Arg Ala Leu Lys Gly Ala Tyr Asp Thr
    290                 295                 300

Trp Lys Leu Lys Lys Arg Leu Glu Lys Arg Lys Ala Phe Pro Tyr Met
305                 310                 315                 320

Pro Asn Trp Asp Asn Asp Tyr Gln Ile Pro Val Gly Leu Thr Gly Leu
                325                 330                 335

Gly Val Phe Thr Leu Glu Val Lys Arg Thr Glu Val Val Val Asp Leu
            340                 345                 350

Lys Glu His Gly Lys Leu Phe Cys Ser His Ser His Tyr Phe Gly Asp
        355                 360                 365

Leu Thr Ala Glu Lys His Pro Ser Arg Tyr His Leu Lys Phe Arg His
    370                 375                 380

Lys Leu Lys Leu Arg Lys Arg Asp Ser Arg Val Glu Pro Thr Ile Gly
385                 390                 395                 400

Pro Trp Ile Glu Ala Ala Leu Arg Glu Ile Thr Ile Gln Lys Lys Pro
                405                 410                 415

Asn Gly Val Phe Tyr Leu Gly Leu Pro Tyr Ala Leu Ser His Gly Ile
            420                 425                 430

Asp Asn Phe Gln Ile Ala Lys Arg Phe Phe Ser Ala Ala Lys Pro Asp
        435                 440                 445

Lys Glu Val Ile Asn Gly Leu Pro Ser Glu Met Val Val Gly Ala Ala
    450                 455                 460

Asp Leu Asn Leu Ser Asn Ile Val Ala Pro Val Lys Ala Arg Ile Gly
465                 470                 475                 480
```

```
Lys Gly Leu Glu Gly Pro Leu His Ala Leu Asp Tyr Gly Tyr Gly Glu
                485                 490                 495

Leu Ile Asp Gly Pro Lys Ile Leu Thr Pro Asp Gly Pro Arg Cys Gly
            500                 505                 510

Glu Leu Ile Ser Leu Lys Arg Asp Ile Val Glu Ile Lys Ser Ala Ile
        515                 520                 525

Lys Glu Phe Lys Ala Cys Gln Arg Glu Gly Leu Thr Met Ser Glu Glu
    530                 535                 540

Thr Thr Thr Trp Leu Ser Glu Val Glu Ser Pro Ser Asp Ser Pro Arg
545                 550                 555                 560

Cys Met Ile Gln Ser Arg Ile Ala Asp Thr Ser Arg Arg Leu Asn Ser
                565                 570                 575

Phe Lys Tyr Gln Met Asn Lys Glu Gly Tyr Gln Asp Leu Ala Glu Ala
            580                 585                 590

Leu Arg Leu Leu Asp Ala Met Asp Ser Tyr Asn Ser Leu Leu Glu Ser
        595                 600                 605

Tyr Gln Arg Met His Leu Ser Pro Gly Glu Gln Ser Pro Lys Glu Ala
    610                 615                 620

Lys Phe Asp Thr Lys Arg Ala Ser Phe Arg Asp Leu Leu Arg Arg Arg
625                 630                 635                 640

Val Ala His Thr Ile Val Glu Tyr Phe Asp Asp Cys Asp Ile Val Phe
                645                 650                 655

Phe Glu Asp Leu Asp Gly Pro Ser Asp Ser Asp Ser Arg Asn Asn Ala
            660                 665                 670

Leu Val Lys Leu Leu Ser Pro Arg Thr Leu Leu Leu Tyr Ile Arg Gln
        675                 680                 685

Ala Leu Glu Lys Arg Gly Ile Gly Met Val Glu Val Ala Lys Asp Gly
    690                 695                 700
```

```
Thr Ser Gln Asn Asn Pro Ile Ser Gly His Val Gly Trp Arg Asn Lys
705                 710                 715                 720

Gln Asn Lys Ser Glu Ile Tyr Phe Tyr Glu Asp Lys Glu Leu Leu Val
                725                 730                 735

Met Asp Ala Asp Glu Val Gly Ala Met Asn Ile Leu Cys Arg Gly Leu
            740                 745                 750

Asn His Ser Val Cys Pro Tyr Ser Phe Val Thr Lys Ala Pro Glu Lys
        755                 760                 765

Lys Asn Asp Glu Lys Lys Glu Gly Asp Tyr Gly Lys Arg Val Lys Arg
    770                 775                 780

Phe Leu Lys Asp Arg Tyr Gly Ser Ser Asn Val Arg Phe Leu Val Ala
785                 790                 795                 800

Ser Met Gly Phe Val Thr Val Thr Thr Lys Arg Pro Lys Asp Ala Leu
                805                 810                 815

Val Gly Lys Arg Leu Tyr Tyr His Gly Gly Glu Leu Val Thr His Asp
            820                 825                 830

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

Val Leu Ala Arg Arg Val Ser Leu Ser Asp Ser Thr Ile Lys Ser Tyr
    850                 855                 860

Lys Ser Phe Ala His Val
865                 870
```

<210> 244
<211> 1093
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Cas12i1 polypeptide sequence

<400> 244

```
Met Ser Asn Lys Glu Lys Asn Ala Ser Glu Thr Arg Lys Ala Tyr Thr
1               5                  10                  15

Thr Lys Met Ile Pro Arg Ser His Asp Arg Met Lys Leu Leu Gly Asn
            20                  25                  30

Phe Met Asp Tyr Leu Met Asp Gly Thr Pro Ile Phe Phe Glu Leu Trp
        35                  40                  45

Asn Gln Phe Gly Gly Gly Ile Asp Arg Asp Ile Ile Ser Gly Thr Ala
    50                  55                  60

Asn Lys Asp Lys Ile Ser Asp Asp Leu Leu Leu Ala Val Asn Trp Phe
65                  70                  75                  80

Lys Val Met Pro Ile Asn Ser Lys Pro Gln Gly Val Ser Pro Ser Asn
                85                  90                  95

Leu Ala Asn Leu Phe Gln Gln Tyr Ser Gly Ser Glu Pro Asp Ile Gln
            100                 105                 110

Ala Gln Glu Tyr Phe Ala Ser Asn Phe Asp Thr Glu Lys His Gln Trp
        115                 120                 125

Lys Asp Met Arg Val Glu Tyr Glu Arg Leu Leu Ala Glu Leu Gln Leu
    130                 135                 140

Ser Arg Ser Asp Met His His Asp Leu Lys Leu Met Tyr Lys Glu Lys
145                 150                 155                 160

Cys Ile Gly Leu Ser Leu Ser Thr Ala His Tyr Ile Thr Ser Val Met
                165                 170                 175

Phe Gly Thr Gly Ala Lys Asn Asn Arg Gln Thr Lys His Gln Phe Tyr
            180                 185                 190

Ser Lys Val Ile Gln Leu Leu Glu Glu Ser Thr Gln Ile Asn Ser Val
        195                 200                 205
```

```
Glu Gln Leu Ala Ser Ile Ile Leu Lys Ala Gly Asp Cys Asp Ser Tyr
    210                 215                 220

Arg Lys Leu Arg Ile Arg Cys Ser Arg Lys Gly Ala Thr Pro Ser Ile
225                 230                 235                 240

Leu Lys Ile Val Gln Asp Tyr Glu Leu Gly Thr Asn His Asp Asp Glu
                245                 250                 255

Val Asn Val Pro Ser Leu Ile Ala Asn Leu Lys Glu Lys Leu Gly Arg
            260                 265                 270

Phe Glu Tyr Glu Cys Glu Trp Lys Cys Met Glu Lys Ile Lys Ala Phe
        275                 280                 285

Leu Ala Ser Lys Val Gly Pro Tyr Tyr Leu Gly Ser Tyr Ser Ala Met
    290                 295                 300

Leu Glu Asn Ala Leu Ser Pro Ile Lys Gly Met Thr Thr Lys Asn Cys
305                 310                 315                 320

Lys Phe Val Leu Lys Gln Ile Asp Ala Lys Asn Asp Ile Lys Tyr Glu
                325                 330                 335

Asn Glu Pro Phe Gly Lys Ile Val Glu Gly Phe Phe Asp Ser Pro Tyr
            340                 345                 350

Phe Glu Ser Asp Thr Asn Val Lys Trp Val Leu His Pro His His Ile
        355                 360                 365

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

Ser Lys Tyr Glu Glu Asp Ile Ala Ser Leu Ser Glu Asp Lys Lys Glu
385                 390                 395                 400

Lys Arg Ile Lys Val Tyr Gln Gly Asp Val Cys Gln Thr Ile Asn Thr
                405                 410                 415

Tyr Cys Glu Glu Val Gly Lys Glu Ala Lys Thr Pro Leu Val Gln Leu
            420                 425                 430
```

```
Leu Arg Tyr Leu Tyr Ser Arg Lys Asp Asp Ile Ala Val Asp Lys Ile
        435                 440                 445

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

Lys Ile Asn Pro Val Ile Gln Lys Tyr Pro Ser Phe Asn Phe Gly Asn
465                 470                 475                 480

Asn Ser Lys Leu Leu Gly Lys Ile Ile Ser Pro Lys Asp Lys Leu Lys
                485                 490                 495

His Asn Leu Lys Cys Asn Arg Asn Gln Val Asp Asn Tyr Ile Trp Ile
            500                 505                 510

Glu Ile Lys Val Leu Asn Thr Lys Thr Met Arg Trp Glu Lys His His
        515                 520                 525

Tyr Ala Leu Ser Ser Thr Arg Phe Leu Glu Glu Val Tyr Tyr Pro Ala
    530                 535                 540

Thr Ser Glu Asn Pro Pro Asp Ala Leu Ala Ala Arg Phe Arg Thr Lys
545                 550                 555                 560

Thr Asn Gly Tyr Glu Gly Lys Pro Ala Leu Ser Ala Glu Gln Ile Glu
                565                 570                 575

Gln Ile Arg Ser Ala Pro Val Gly Leu Arg Lys Val Lys Lys Arg Gln
            580                 585                 590

Met Arg Leu Glu Ala Ala Arg Gln Gln Asn Leu Leu Pro Arg Tyr Thr
        595                 600                 605

Trp Gly Lys Asp Phe Asn Ile Asn Ile Cys Lys Arg Gly Asn Asn Phe
    610                 615                 620

Glu Val Thr Leu Ala Thr Lys Val Lys Lys Lys Lys Glu Lys Asn Tyr
625                 630                 635                 640
```

```
Lys Val Val Leu Gly Tyr Asp Ala Asn Ile Val Arg Lys Asn Thr Tyr
                645                 650                 655

Ala Ala Ile Glu Ala His Ala Asn Gly Asp Gly Val Ile Asp Tyr Asn
            660                 665                 670

Asp Leu Pro Val Lys Pro Ile Glu Ser Gly Phe Val Thr Val Glu Ser
        675                 680                 685

Gln Val Arg Asp Lys Ser Tyr Asp Gln Leu Ser Tyr Asn Gly Val Lys
    690                 695                 700

Leu Leu Tyr Cys Lys Pro His Val Glu Ser Arg Arg Ser Phe Leu Glu
705                 710                 715                 720

Lys Tyr Arg Asn Gly Thr Met Lys Asp Asn Arg Gly Asn Asn Ile Gln
                725                 730                 735

Ile Asp Phe Met Lys Asp Phe Glu Ala Ile Ala Asp Asp Glu Thr Ser
            740                 745                 750

Leu Tyr Tyr Phe Asn Met Lys Tyr Cys Lys Leu Leu Gln Ser Ser Ile
        755                 760                 765

Arg Asn His Ser Ser Gln Ala Lys Glu Tyr Arg Glu Glu Ile Phe Glu
    770                 775                 780

Leu Leu Arg Asp Gly Lys Leu Ser Val Leu Lys Leu Ser Ser Leu Ser
785                 790                 795                 800

Asn Leu Ser Phe Val Met Phe Lys Val Ala Lys Ser Leu Ile Gly Thr
                805                 810                 815

Tyr Phe Gly His Leu Leu Lys Lys Pro Lys Asn Ser Lys Ser Asp Val
            820                 825                 830

Lys Ala Pro Pro Ile Thr Asp Glu Asp Lys Gln Lys Ala Asp Pro Glu
        835                 840                 845

Met Phe Ala Leu Arg Leu Ala Leu Glu Glu Lys Arg Leu Asn Lys Val
    850                 855                 860
```

```
Lys Ser Lys Lys Glu Val Ile Ala Asn Lys Ile Val Ala Lys Ala Leu
865                 870                 875                 880

Glu Leu Arg Asp Lys Tyr Gly Pro Val Leu Ile Lys Gly Glu Asn Ile
                885                 890                 895

Ser Asp Thr Thr Lys Lys Gly Lys Lys Ser Ser Thr Asn Ser Phe Leu
            900                 905                 910

Met Asp Trp Leu Ala Arg Gly Val Ala Asn Lys Val Lys Glu Met Val
        915                 920                 925

Met Met His Gln Gly Leu Glu Phe Val Glu Val Asn Pro Asn Phe Thr
    930                 935                 940

Ser His Gln Asp Pro Phe Val His Lys Asn Pro Glu Asn Thr Phe Arg
945                 950                 955                 960

Ala Arg Tyr Ser Arg Cys Thr Pro Ser Glu Leu Thr Glu Lys Asn Arg
                965                 970                 975

Lys Glu Ile Leu Ser Phe Leu Ser Asp Lys Pro Ser Lys Arg Pro Thr
            980                 985                 990

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

Leu Lys  Lys Asn Asp Val Leu  Gly Val Ser Leu Glu  Lys Phe Lys
    1010                1015                1020

Gln Ile  Met Ala Asn Ile Leu  His Gln Arg Ser Glu  Asp Gln Leu
    1025                1030                1035

Leu Phe  Pro Ser Arg Gly Gly  Met Phe Tyr Leu Ala  Thr Tyr Lys
    1040                1045                1050

Leu Asp  Ala Asp Ala Thr Ser  Val Asn Trp Asn Gly  Lys Gln Phe
    1055                1060                1065
```

Trp Val Cys Asn Ala Asp Leu Val Ala Ala Tyr Asn Val Gly Leu
    1070   1075   1080

Val Asp Ile Gln Lys Asp Phe Lys Lys Lys
    1085   1090

<210> 245
<211> 1054
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: Cas12i2 polypeptide sequence

<400> 245

Met Ser Ser Ala Ile Lys Ser Tyr Lys Ser Val Leu Arg Pro Asn Glu
1   5   10   15

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

Ser Ala Phe Phe Phe Lys Met Leu Gln Gly Leu Phe Gly Gly Ile Thr
        35   40   45

Pro Glu Ile Val Arg Phe Ser Thr Glu Gln Glu Lys Gln Gln Gln Asp
    50   55   60

Ile Ala Leu Trp Cys Ala Val Asn Trp Phe Arg Pro Val Ser Gln Asp
65   70   75   80

Ser Leu Thr His Thr Ile Ala Ser Asp Asn Leu Val Glu Lys Phe Glu
                85   90   95

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

Ala Ser Ile Gly Glu Ser Tyr Tyr Trp Asn Asp Cys Arg Gln Gln Tyr
        115   120   125

Tyr Asp Leu Cys Arg Glu Leu Gly Val Glu Val Ser Asp Leu Thr His
    130   135   140

```
Asp Leu Glu Ile Leu Cys Arg Glu Lys Cys Leu Ala Val Ala Thr Glu
145                 150                 155                 160

Ser Asn Gln Asn Asn Ser Ile Ile Ser Val Leu Phe Gly Thr Gly Glu
                165                 170                 175

Lys Glu Asp Arg Ser Val Lys Leu Arg Ile Thr Lys Lys Ile Leu Glu
            180                 185                 190

Ala Ile Ser Asn Leu Lys Glu Ile Pro Lys Asn Val Ala Pro Ile Gln
        195                 200                 205

Glu Ile Ile Leu Asn Val Ala Lys Ala Thr Lys Glu Thr Phe Arg Gln
    210                 215                 220

Val Tyr Ala Gly Asn Leu Gly Ala Pro Ser Thr Leu Glu Lys Phe Ile
225                 230                 235                 240

Ala Lys Asp Gly Gln Lys Glu Phe Asp Leu Lys Lys Leu Gln Thr Asp
                245                 250                 255

Leu Lys Lys Val Ile Arg Gly Lys Ser Lys Glu Arg Asp Trp Cys Cys
            260                 265                 270

Gln Glu Glu Leu Arg Ser Tyr Val Glu Gln Asn Thr Ile Gln Tyr Asp
        275                 280                 285

Leu Trp Ala Trp Gly Glu Met Phe Asn Lys Ala His Thr Ala Leu Lys
    290                 295                 300

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

Phe Lys Glu Ile Gln Ser Leu Asn Asn Leu Leu Val Val Lys Lys Leu
                325                 330                 335

Asn Asp Phe Phe Asp Ser Glu Phe Phe Ser Gly Glu Glu Thr Tyr Thr
            340                 345                 350

Ile Cys Val His His Leu Gly Gly Lys Asp Leu Ser Lys Leu Tyr Lys
        355                 360                 365
```

```
Ala Trp Glu Asp Asp Pro Ala Asp Pro Glu Asn Ala Ile Val Val Leu
    370                 375                 380

Cys Asp Asp Leu Lys Asn Asn Phe Lys Lys Glu Pro Ile Arg Asn Ile
385                 390                 395                 400

Leu Arg Tyr Ile Phe Thr Ile Arg Gln Glu Cys Ser Ala Gln Asp Ile
                405                 410                 415

Leu Ala Ala Ala Lys Tyr Asn Gln Gln Leu Asp Arg Tyr Lys Ser Gln
            420                 425                 430

Lys Ala Asn Pro Ser Val Leu Gly Asn Gln Gly Phe Thr Trp Thr Asn
        435                 440                 445

Ala Val Ile Leu Pro Glu Lys Ala Gln Arg Asn Asp Arg Pro Asn Ser
    450                 455                 460

Leu Asp Leu Arg Ile Trp Leu Tyr Leu Lys Leu Arg His Pro Asp Gly
465                 470                 475                 480

Arg Trp Lys Lys His His Ile Pro Phe Tyr Asp Thr Arg Phe Phe Gln
                485                 490                 495

Glu Ile Tyr Ala Ala Gly Asn Ser Pro Val Asp Thr Cys Gln Phe Arg
            500                 505                 510

Thr Pro Arg Phe Gly Tyr His Leu Pro Lys Leu Thr Asp Gln Thr Ala
        515                 520                 525

Ile Arg Val Asn Lys Lys His Val Lys Ala Ala Lys Thr Glu Ala Arg
    530                 535                 540

Ile Arg Leu Ala Ile Gln Gln Gly Thr Leu Pro Val Ser Asn Leu Lys
545                 550                 555                 560

Ile Thr Glu Ile Ser Ala Thr Ile Asn Ser Lys Gly Gln Val Arg Ile
                565                 570                 575
```

```
Pro Val Lys Phe Asp Val Gly Arg Gln Lys Gly Thr Leu Gln Ile Gly
            580                 585                 590

Asp Arg Phe Cys Gly Tyr Asp Gln Asn Gln Thr Ala Ser His Ala Tyr
        595                 600                 605

Ser Leu Trp Glu Val Val Lys Glu Gly Gln Tyr His Lys Glu Leu Gly
    610                 615                 620

Cys Phe Val Arg Phe Ile Ser Ser Gly Asp Ile Val Ser Ile Thr Glu
625                 630                 635                 640

Asn Arg Gly Asn Gln Phe Asp Gln Leu Ser Tyr Glu Gly Leu Ala Tyr
                645                 650                 655

Pro Gln Tyr Ala Asp Trp Arg Lys Lys Ala Ser Lys Phe Val Ser Leu
            660                 665                 670

Trp Gln Ile Thr Lys Lys Asn Lys Lys Lys Glu Ile Val Thr Val Glu
        675                 680                 685

Ala Lys Glu Lys Phe Asp Ala Ile Cys Lys Tyr Gln Pro Arg Leu Tyr
    690                 695                 700

Lys Phe Asn Lys Glu Tyr Ala Tyr Leu Leu Arg Asp Ile Val Arg Gly
705                 710                 715                 720

Lys Ser Leu Val Glu Leu Gln Gln Ile Arg Gln Glu Ile Phe Arg Phe
                725                 730                 735

Ile Glu Gln Asp Cys Gly Val Thr Arg Leu Gly Ser Leu Ser Leu Ser
            740                 745                 750

Thr Leu Glu Thr Val Lys Ala Val Lys Gly Ile Ile Tyr Ser Tyr Phe
        755                 760                 765

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

Arg Lys Glu Phe Asp Pro Glu Leu Phe Ala Leu Leu Glu Lys Leu Glu
785                 790                 795                 800
```

```
Leu Ile Arg Thr Arg Lys Lys Lys Gln Lys Val Glu Arg Ile Ala Asn
                805                 810                 815

Ser Leu Ile Gln Thr Cys Leu Glu Asn Asn Ile Lys Phe Ile Arg Gly
            820                 825                 830

Glu Gly Asp Leu Ser Thr Thr Asn Asn Ala Thr Lys Lys Lys Ala Asn
        835                 840                 845

Ser Arg Ser Met Asp Trp Leu Ala Arg Gly Val Phe Asn Lys Ile Arg
    850                 855                 860

Gln Leu Ala Pro Met His Asn Ile Thr Leu Phe Gly Cys Gly Ser Leu
865                 870                 875                 880

Tyr Thr Ser His Gln Asp Pro Leu Val His Arg Asn Pro Asp Lys Ala
                885                 890                 895

Met Lys Cys Arg Trp Ala Ala Ile Pro Val Lys Asp Ile Gly Asp Trp
            900                 905                 910

Val Leu Arg Lys Leu Ser Gln Asn Leu Arg Ala Lys Asn Ile Gly Thr
        915                 920                 925

Gly Glu Tyr Tyr His Gln Gly Val Lys Glu Phe Leu Ser His Tyr Glu
    930                 935                 940

Leu Gln Asp Leu Glu Glu Glu Leu Leu Lys Trp Arg Ser Asp Arg Lys
945                 950                 955                 960

Ser Asn Ile Pro Cys Trp Val Leu Gln Asn Arg Leu Ala Glu Lys Leu
                965                 970                 975

Gly Asn Lys Glu Ala Val Val Tyr Ile Pro Val Arg Gly Gly Arg Ile
            980                 985                 990

Tyr Phe Ala Thr His Lys Val Ala  Thr Gly Ala Val Ser  Ile Val Phe
        995                 1000                1005
```

Asp Gln Lys Gln Val Trp Val Cys Asn Ala Asp His Val Ala Ala
    1010   1015   1020

Ala Asn Ile Ala Leu Thr Val Lys Gly Ile Gly Glu Gln Ser Ser
    1025   1030   1035

Asp Glu Glu Asn Pro Asp Gly Ser Arg Ile Lys Leu Gln Leu Thr
    1040   1045   1050

Ser

<210> 246
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 246

Gly Gly Gly Ser
1

<210> 247
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 247

Gly Gly Gly Gly Ser
1   5

<210> 248
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 248

Glu Ala Ala Ala Lys
1 5

<210> 249
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 249

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 250
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: linker peptide sequence

<400> 250

Gly Gly Ser Gly Gly Ser
1 5

<210> 251
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 251

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
1 5 10 15

Glu Ser Ser Gly
20

<210> 252
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 252
ggaggctctg gaggaagc 18

<210> 253
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 253
ggctcttctg gatctgaaac acctggcaca agcgagagcg ccacccctga gagctctggc 60

<210> 254
<211> 1129
<212> PRT
<213> Alicyclobacillus acidiphilus

<220>
<221> MISC_FEATURE
<222> (1)..(1129)
<223> AacCas12b polypeptide sequence

<400> 254

Met Ala Val Lys Ser Met Lys Val Lys Leu Arg Leu Asp Asn Met Pro
1 5 10 15

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

Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu
35 40 45

Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Tyr Lys Thr
50 55 60

Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65 70 75 80

```
Val Glu Asn Gly His Cys Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu
                85                  90                  95

Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
            100                 105                 110

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

Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn
    130                 135                 140

Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu
145                 150                 155                 160

Glu Glu Lys Ala Lys Ala Glu Ala Arg Lys Ser Thr Asp Arg Thr Ala
                165                 170                 175

Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg
            180                 185                 190

Val Tyr Thr Asp Ser Asp Met Ser Ser Val Gln Trp Lys Pro Leu Arg
        195                 200                 205

Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala
    210                 215                 220

Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Glu
225                 230                 235                 240

Ala Tyr Ala Lys Leu Val Glu Gln Lys Ser Arg Phe Glu Gln Lys Asn
                245                 250                 255

Phe Val Gly Gln Glu His Leu Val Gln Leu Val Asn Gln Leu Gln Gln
            260                 265                 270

Asp Met Lys Glu Ala Ser His Gly Leu Glu Ser Lys Glu Gln Thr Ala
        275                 280                 285

His Tyr Leu Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu
```

```
    290                 295                 300

Lys Trp Glu Lys Leu Asp Pro Asp Ala Pro Phe Asp Leu Tyr Asp Thr
305                 310                 315                 320

Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His
                325                 330                 335

Asp Leu Phe Ala Lys Leu Ala Glu Pro Lys Tyr Gln Ala Leu Trp Arg
            340                 345                 350

Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Val
        355                 360                 365

Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
    370                 375                 380

Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
385                 390                 395                 400

Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Gly Arg His
                405                 410                 415

Ala Ile Arg Phe Gln Lys Leu Leu Thr Val Glu Asp Gly Val Ala Lys
            420                 425                 430

Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Ala Gln Leu Asp
        435                 440                 445

Asp Leu Leu Pro Arg Asp Pro His Glu Leu Val Ala Leu Tyr Phe Gln
    450                 455                 460

Asp Tyr Gly Ala Glu Gln His Leu Ala Gly Glu Phe Gly Gly Ala Lys
465                 470                 475                 480

Ile Gln Tyr Arg Arg Asp Gln Leu Asn His Leu His Ala Arg Arg Gly
                485                 490                 495

Ala Arg Asp Val Tyr Leu Asn Leu Ser Val Arg Val Gln Ser Gln Ser
            500                 505                 510
```

```
Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
        515                 520                 525

Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp
    530                 535                 540

Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu
545                 550                 555                 560

Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser
                565                 570                 575

Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
            580                 585                 590

Asn Ser Glu Gly Arg Val Pro Phe Cys Phe Pro Ile Glu Gly Asn Glu
        595                 600                 605

Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly
    610                 615                 620

Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg
625                 630                 635                 640

Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val
                645                 650                 655

Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys
            660                 665                 670

Leu Ile Glu Gln Pro Met Asp Ala Asn Gln Met Thr Pro Asp Trp Arg
        675                 680                 685

Glu Ala Phe Glu Asp Glu Leu Gln Lys Leu Lys Ser Leu Tyr Gly Ile
    690                 695                 700

Cys Gly Asp Arg Glu Trp Thr Glu Ala Val Tyr Glu Ser Val Arg Arg
705                 710                 715                 720

Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp Val
```

```
                725                 730                 735

Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Gln Lys Asp Val Val
            740                 745                 750

Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr Lys
        755                 760                 765

Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val Ile
    770                 775                 780

Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His Ile
785                 790                 795                 800

Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile Ile
                805                 810                 815

Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Asp Glu Arg Gly Lys
            820                 825                 830

Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu
        835                 840                 845

Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn
    850                 855                 860

Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Leu
865                 870                 875                 880

Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala
                885                 890                 895

Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys
            900                 905                 910

Arg Arg Val Pro Ala Arg Cys Ala Arg Glu Gln Asn Pro Glu Pro Phe
        915                 920                 925

Pro Trp Trp Leu Asn Lys Phe Val Ala Glu His Lys Leu Asp Gly Cys
    930                 935                 940
```

```
Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Phe Phe
945                 950                 955                 960


Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala
                965                 970                 975


Asp Leu Asn Ala Ala Gln Asn Leu Gln Arg Arg Leu Trp Ser Asp Phe
            980                 985                 990


Asp Ile Ser Gln Ile Arg Leu Arg  Cys Asp Trp Gly Glu  Val Asp Gly
        995                 1000                1005


Glu Pro  Val Leu Ile Pro Arg  Thr Thr Gly Lys Arg  Thr Ala Asp
    1010                1015                1020


Ser Tyr  Gly Asn Lys Val Phe  Tyr Thr Lys Thr Gly  Val Thr Tyr
    1025                1030                1035


Tyr Glu  Arg Glu Arg Gly Lys  Lys Arg Arg Lys Val  Phe Ala Gln
    1040                1045                1050


Glu Glu  Leu Ser Glu Glu Glu  Ala Glu Leu Leu Val  Glu Ala Asp
    1055                1060                1065


Glu Ala  Arg Glu Lys Ser Val  Val Leu Met Arg Asp  Pro Ser Gly
    1070                1075                1080


Ile Ile  Asn Arg Gly Asp Trp  Thr Arg Gln Lys Glu  Phe Trp Ser
    1085                1090                1095


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


Ser Arg  Val Arg Leu Gln Glu  Ser Ala Cys Glu Asn  Thr Gly Asp
    1115                1120                1125


Ile
```

<210> 255

<211> 1140
<212> PRT
<213> Bacillus hisashii

<220>
<221> MISC_FEATURE
<222> (1)..(1140)
<223> BhCas12b polypeptide sequence

<400> 255

```
Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1               5                   10                  15


Ala Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
            20                  25                  30


Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
        35                  40                  45


Tyr Tyr Met Asn Ile Leu Lys Leu Ile Arg Gln Glu Ala Ile Tyr Glu
    50                  55                  60


His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
65                  70                  75                  80


Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
                85                  90                  95


Ser Phe Thr His Glu Val Asp Lys Asp Glu Val Phe Asn Ile Leu Arg
            100                 105                 110


Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
        115                 120                 125


Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
    130                 135                 140


Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
145                 150                 155                 160


Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
                165                 170                 175
```

```
Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
            180                 185                 190

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Tyr Thr Asp
        195                 200                 205

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
    210                 215                 220

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
225                 230                 235                 240

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
                245                 250                 255

Tyr Glu Lys Val Glu Lys Glu Tyr Lys Thr Leu Glu Glu Arg Ile Lys
            260                 265                 270

Glu Asp Ile Gln Ala Leu Lys Ala Leu Glu Gln Tyr Glu Lys Glu Arg
        275                 280                 285

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
    290                 295                 300

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
305                 310                 315                 320

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
                325                 330                 335

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
            340                 345                 350

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
        355                 360                 365

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
    370                 375                 380
```

```
Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
385                 390                 395                 400

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
                405                 410                 415

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
            420                 425                 430

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
        435                 440                 445

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
    450                 455                 460

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

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
                485                 490                 495

Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
            500                 505                 510

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
        515                 520                 525

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
    530                 535                 540

Ile His Arg Asp Asp Phe Pro Lys Val Val Asn Phe Lys Pro Lys Glu
545                 550                 555                 560

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
                565                 570                 575

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
            580                 585                 590

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
        595                 600                 605
```

```
Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
    610                 615                 620

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
625                 630                 635                 640

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
                645                 650                 655

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
            660                 665                 670

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
        675                 680                 685

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
    690                 695                 700

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
705                 710                 715                 720

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
                725                 730                 735

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
            740                 745                 750

Tyr Gly Ile Ser Leu Lys Asn Ile Asp Glu Ile Asp Arg Thr Arg Lys
        755                 760                 765

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
    770                 775                 780

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
785                 790                 795                 800

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
                805                 810                 815
```

```
Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
            820                 825                 830

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
        835                 840                 845

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
    850                 855                 860

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
865                 870                 875                 880

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
                885                 890                 895

Phe His Ala Lys Thr Gly Ser Pro Gly Ile Arg Cys Ser Val Val Thr
            900                 905                 910

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
        915                 920                 925

Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
    930                 935                 940

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
945                 950                 955                 960

Lys Cys Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
                965                 970                 975

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
            980                 985                 990

Ala Tyr Gln Val Asp Gly Gln Thr  Val Tyr Ile Pro Glu  Ser Lys Asp
        995                 1000                1005

Gln Lys  Gln Lys Ile Ile Glu  Glu Phe Gly Glu Gly  Tyr Phe Ile
    1010                1015                1020

Leu Lys  Asp Gly Val Tyr Glu  Trp Val Asn Ala Gly  Lys Leu Lys
    1025                1030                1035
```

```
Ile Lys  Lys Gly Ser Ser Lys  Gln Ser Ser Ser Glu  Leu Val Asp
    1040                1045                1050


Ser Asp  Ile Leu Lys Asp Ser  Phe Asp Leu Ala Ser  Glu Leu Lys
    1055                1060                1065


Gly Glu  Lys Leu Met Leu Tyr  Arg Asp Pro Ser Gly  Asn Val Phe
    1070                1075                1080


Pro Ser  Asp Lys Trp Met Ala  Ala Gly Val Phe Phe  Gly Lys Leu
    1085                1090                1095


Glu Arg  Ile Leu Ile Ser Lys  Leu Thr Asn Gln Tyr  Ser Ile Ser
    1100                1105                1110


Thr Ile  Glu Asp Asp Ser Ser  Lys Gln Ser Met Lys  Arg Pro Ala
    1115                1120                1125


Ala Thr  Lys Lys Ala Gly Gln  Ala Lys Lys Lys Lys
    1130                1135                1140


<210>  256
<211>  1108
<212>  PRT
<213>  Bacillus thermoamylovorans


<220>
<221>  MISC_FEATURE
<222>  (1)..(1108)
<223>  BTCas12b.BTCas12b polypeptide sequence

<400>  256

Met Ala Thr Arg Ser Phe Ile Leu Lys Ile Glu Pro Asn Glu Glu Val
1               5                   10                  15


Lys Lys Gly Leu Trp Lys Thr His Glu Val Leu Asn His Gly Ile Ala
            20                  25                  30


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

```
His His Glu Gln Asp Pro Lys Asn Pro Lys Lys Val Ser Lys Ala Glu
    50                  55                  60

Ile Gln Ala Glu Leu Trp Asp Phe Val Leu Lys Met Gln Lys Cys Asn
65                  70                  75                  80

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

Glu Leu Tyr Glu Glu Leu Val Pro Ser Ser Val Glu Lys Lys Gly Glu
            100                 105                 110

Ala Asn Gln Leu Ser Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn
        115                 120                 125

Ser Gln Ser Gly Lys Gly Thr Ala Ser Ser Gly Arg Lys Pro Arg Trp
    130                 135                 140

Tyr Asn Leu Lys Ile Ala Gly Asp Pro Ser Trp Glu Glu Glu Lys Lys
145                 150                 155                 160

Lys Trp Glu Glu Asp Lys Lys Lys Asp Pro Leu Ala Lys Ile Leu Gly
                165                 170                 175

Lys Leu Ala Glu Tyr Gly Leu Ile Pro Leu Phe Ile Pro Phe Thr Asp
            180                 185                 190

Ser Asn Glu Pro Ile Val Lys Glu Ile Lys Trp Met Glu Lys Ser Arg
        195                 200                 205

Asn Gln Ser Val Arg Arg Leu Asp Lys Asp Met Phe Ile Gln Ala Leu
    210                 215                 220

Glu Arg Phe Leu Ser Trp Glu Ser Trp Asn Leu Lys Val Lys Glu Glu
225                 230                 235                 240

Tyr Glu Lys Val Glu Lys Glu His Lys Thr Leu Glu Glu Arg Ile Lys
                245                 250                 255

Glu Asp Ile Gln Ala Phe Lys Ser Leu Glu Gln Tyr Glu Lys Glu Arg
```

```
            260                 265                 270

Gln Glu Gln Leu Leu Arg Asp Thr Leu Asn Thr Asn Glu Tyr Arg Leu
        275                 280                 285

Ser Lys Arg Gly Leu Arg Gly Trp Arg Glu Ile Ile Gln Lys Trp Leu
    290                 295                 300

Lys Met Asp Glu Asn Glu Pro Ser Glu Lys Tyr Leu Glu Val Phe Lys
305                 310                 315                 320

Asp Tyr Gln Arg Lys His Pro Arg Glu Ala Gly Asp Tyr Ser Val Tyr
                325                 330                 335

Glu Phe Leu Ser Lys Lys Glu Asn His Phe Ile Trp Arg Asn His Pro
            340                 345                 350

Glu Tyr Pro Tyr Leu Tyr Ala Thr Phe Cys Glu Ile Asp Lys Lys Lys
        355                 360                 365

Lys Asp Ala Lys Gln Gln Ala Thr Phe Thr Leu Ala Asp Pro Ile Asn
    370                 375                 380

His Pro Leu Trp Val Arg Phe Glu Glu Arg Ser Gly Ser Asn Leu Asn
385                 390                 395                 400

Lys Tyr Arg Ile Leu Thr Glu Gln Leu His Thr Glu Lys Leu Lys Lys
                405                 410                 415

Lys Leu Thr Val Gln Leu Asp Arg Leu Ile Tyr Pro Thr Glu Ser Gly
            420                 425                 430

Gly Trp Glu Glu Lys Gly Lys Val Asp Ile Val Leu Leu Pro Ser Arg
        435                 440                 445

Gln Phe Tyr Asn Gln Ile Phe Leu Asp Ile Glu Glu Lys Gly Lys His
    450                 455                 460

Ala Phe Thr Tyr Lys Asp Glu Ser Ile Lys Phe Pro Leu Lys Gly Thr
465                 470                 475                 480
```

```
Leu Gly Gly Ala Arg Val Gln Phe Asp Arg Asp His Leu Arg Arg Tyr
                485                 490                 495

Pro His Lys Val Glu Ser Gly Asn Val Gly Arg Ile Tyr Phe Asn Met
            500                 505                 510

Thr Val Asn Ile Glu Pro Thr Glu Ser Pro Val Ser Lys Ser Leu Lys
        515                 520                 525

Ile His Arg Asp Asp Phe Pro Lys Phe Val Asn Phe Lys Pro Lys Glu
    530                 535                 540

Leu Thr Glu Trp Ile Lys Asp Ser Lys Gly Lys Lys Leu Lys Ser Gly
545                 550                 555                 560

Ile Glu Ser Leu Glu Ile Gly Leu Arg Val Met Ser Ile Asp Leu Gly
                565                 570                 575

Gln Arg Gln Ala Ala Ala Ala Ser Ile Phe Glu Val Val Asp Gln Lys
            580                 585                 590

Pro Asp Ile Glu Gly Lys Leu Phe Phe Pro Ile Lys Gly Thr Glu Leu
        595                 600                 605

Tyr Ala Val His Arg Ala Ser Phe Asn Ile Lys Leu Pro Gly Glu Thr
    610                 615                 620

Leu Val Lys Ser Arg Glu Val Leu Arg Lys Ala Arg Glu Asp Asn Leu
625                 630                 635                 640

Lys Leu Met Asn Gln Lys Leu Asn Phe Leu Arg Asn Val Leu His Phe
                645                 650                 655

Gln Gln Phe Glu Asp Ile Thr Glu Arg Glu Lys Arg Val Thr Lys Trp
            660                 665                 670

Ile Ser Arg Gln Glu Asn Ser Asp Val Pro Leu Val Tyr Gln Asp Glu
        675                 680                 685

Leu Ile Gln Ile Arg Glu Leu Met Tyr Lys Pro Tyr Lys Asp Trp Val
```

```
    690                 695                 700

Ala Phe Leu Lys Gln Leu His Lys Arg Leu Glu Val Glu Ile Gly Lys
705                 710                 715                 720

Glu Val Lys His Trp Arg Lys Ser Leu Ser Asp Gly Arg Lys Gly Leu
                725                 730                 735

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

Phe Leu Leu Arg Trp Ser Leu Arg Pro Thr Glu Pro Gly Glu Val Arg
        755                 760                 765

Arg Leu Glu Pro Gly Gln Arg Phe Ala Ile Asp Gln Leu Asn His Leu
    770                 775                 780

Asn Ala Leu Lys Glu Asp Arg Leu Lys Lys Met Ala Asn Thr Ile Ile
785                 790                 795                 800

Met His Ala Leu Gly Tyr Cys Tyr Asp Val Arg Lys Lys Lys Trp Gln
                805                 810                 815

Ala Lys Asn Pro Ala Cys Gln Ile Ile Leu Phe Glu Asp Leu Ser Asn
            820                 825                 830

Tyr Asn Pro Tyr Glu Glu Arg Ser Arg Phe Glu Asn Ser Lys Leu Met
        835                 840                 845

Lys Trp Ser Arg Arg Glu Ile Pro Arg Gln Val Ala Leu Gln Gly Glu
    850                 855                 860

Ile Tyr Gly Leu Gln Val Gly Glu Val Gly Ala Gln Phe Ser Ser Arg
865                 870                 875                 880

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

Lys Glu Lys Leu Gln Asp Asn Arg Phe Phe Lys Asn Leu Gln Arg Glu
            900                 905                 910
```

```
Gly Arg Leu Thr Leu Asp Lys Ile Ala Val Leu Lys Glu Gly Asp Leu
        915                 920                 925

Tyr Pro Asp Lys Gly Gly Glu Lys Phe Ile Ser Leu Ser Lys Asp Arg
    930                 935                 940

Lys Leu Val Thr Thr His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln
945                 950                 955                 960

Lys Arg Phe Trp Thr Arg Thr His Gly Phe Tyr Lys Val Tyr Cys Lys
                965                 970                 975

Ala Tyr Gln Val Asp Gly Gln Thr Val Tyr Ile Pro Glu Ser Lys Asp
            980                 985                 990

Gln Lys Gln Lys Ile Ile Glu Glu  Phe Gly Glu Gly Tyr  Phe Ile Leu
        995                 1000                1005

Lys Asp  Gly Val Tyr Glu Trp  Gly Asn Ala Gly Lys  Leu Lys Ile
    1010                1015                1020

Lys Lys  Gly Ser Ser Lys Gln  Ser Ser Ser Glu Leu  Val Asp Ser
    1025                1030                1035

Asp Ile  Leu Lys Asp Ser Phe  Asp Leu Ala Ser Glu  Leu Lys Gly
    1040                1045                1050

Glu Lys  Leu Met Leu Tyr Arg  Asp Pro Ser Gly Asn  Val Phe Pro
    1055                1060                1065

Ser Asp  Lys Trp Met Ala Ala  Gly Val Phe Phe Gly  Lys Leu Glu
    1070                1075                1080

Arg Ile  Leu Ile Ser Lys Leu  Thr Asn Gln Tyr Ser  Ile Ser Thr
    1085                1090                1095

Ile Glu  Asp Asp Ser Ser Lys  Gln Ser Met
    1100                1105
```

<210> 257

<211> 1112
<212> PRT
<213> Bacillus sp. V3-13

<220>
<221> MISC_FEATURE
<222> (1)..(1112)
<223> BvCas12b (Bacillus sp. V3-13) polypeptide sequence

<400> 257

Met Ala Ile Arg Ser Ile Lys Leu Lys Met Lys Thr Asn Ser Gly Thr
1 5 10 15

Asp Ser Ile Tyr Leu Arg Lys Ala Leu Trp Arg Thr His Gln Leu Ile
20 25 30

Asn Glu Gly Ile Ala Tyr Tyr Met Asn Leu Leu Thr Leu Tyr Arg Gln
35 40 45

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

Asn Ile Ile Arg Asn Gln Gln Arg Asn Asn Gly Ser Ser Glu Glu His
65 70 75 80

Gly Ser Asp Gln Glu Ile Leu Ala Leu Leu Arg Gln Leu Tyr Glu Leu
85 90 95

Ile Ile Pro Ser Ser Ile Gly Glu Ser Gly Asp Ala Asn Gln Leu Gly
100 105 110

Asn Lys Phe Leu Tyr Pro Leu Val Asp Pro Asn Ser Gln Ser Gly Lys
115 120 125

Gly Thr Ser Asn Ala Gly Arg Lys Pro Arg Trp Lys Arg Leu Lys Glu
130 135 140

Glu Gly Asn Pro Asp Trp Glu Leu Glu Lys Lys Lys Asp Glu Glu Arg
145 150 155 160

Lys Ala Lys Asp Pro Thr Val Lys Ile Phe Asp Asn Leu Asn Lys Tyr
165 170 175

```
Gly Leu Leu Pro Leu Phe Pro Leu Phe Thr Asn Ile Gln Lys Asp Ile
            180                 185                 190

Glu Trp Leu Pro Leu Gly Lys Arg Gln Ser Val Arg Lys Trp Asp Lys
        195                 200                 205

Asp Met Phe Ile Gln Ala Ile Glu Arg Leu Leu Ser Trp Glu Ser Trp
    210                 215                 220

Asn Arg Arg Val Ala Asp Glu Tyr Lys Gln Leu Lys Glu Lys Thr Glu
225                 230                 235                 240

Ser Tyr Tyr Lys Glu His Leu Thr Gly Gly Glu Glu Trp Ile Glu Lys
                245                 250                 255

Ile Arg Lys Phe Glu Lys Glu Arg Asn Met Glu Leu Glu Lys Asn Ala
            260                 265                 270

Phe Ala Pro Asn Asp Gly Tyr Phe Ile Thr Ser Arg Gln Ile Arg Gly
        275                 280                 285

Trp Asp Arg Val Tyr Glu Lys Trp Ser Lys Leu Pro Glu Ser Ala Ser
    290                 295                 300

Pro Glu Glu Leu Trp Lys Val Val Ala Glu Gln Gln Asn Lys Met Ser
305                 310                 315                 320

Glu Gly Phe Gly Asp Pro Lys Val Phe Ser Phe Leu Ala Asn Arg Glu
                325                 330                 335

Asn Arg Asp Ile Trp Arg Gly His Ser Glu Arg Ile Tyr His Ile Ala
            340                 345                 350

Ala Tyr Asn Gly Leu Gln Lys Lys Leu Ser Arg Thr Lys Glu Gln Ala
        355                 360                 365

Thr Phe Thr Leu Pro Asp Ala Ile Glu His Pro Leu Trp Ile Arg Tyr
    370                 375                 380
```

```
Glu Ser Pro Gly Gly Thr Asn Leu Asn Leu Phe Lys Leu Glu Glu Lys
385                 390                 395                 400

Gln Lys Lys Asn Tyr Tyr Val Thr Leu Ser Lys Ile Ile Trp Pro Ser
                405                 410                 415

Glu Glu Lys Trp Ile Glu Lys Glu Asn Ile Glu Ile Pro Leu Ala Pro
            420                 425                 430

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

Lys Gln Glu Ile Ser Phe Ser Asp Tyr Ser Ser Arg Ile Ser Leu Asp
    450                 455                 460

Gly Val Leu Gly Gly Ser Arg Ile Gln Phe Asn Arg Lys Tyr Ile Lys
465                 470                 475                 480

Asn His Lys Glu Leu Leu Gly Glu Gly Asp Ile Gly Pro Val Phe Phe
                485                 490                 495

Asn Leu Val Val Asp Val Ala Pro Leu Gln Glu Thr Arg Asn Gly Arg
            500                 505                 510

Leu Gln Ser Pro Ile Gly Lys Ala Leu Lys Val Ile Ser Ser Asp Phe
        515                 520                 525

Ser Lys Val Ile Asp Tyr Lys Pro Lys Glu Leu Met Asp Trp Met Asn
    530                 535                 540

Thr Gly Ser Ala Ser Asn Ser Phe Gly Val Ala Ser Leu Leu Glu Gly
545                 550                 555                 560

Met Arg Val Met Ser Ile Asp Met Gly Gln Arg Thr Ser Ala Ser Val
                565                 570                 575

Ser Ile Phe Glu Val Val Lys Glu Leu Pro Lys Asp Gln Glu Gln Lys
            580                 585                 590

Leu Phe Tyr Ser Ile Asn Asp Thr Glu Leu Phe Ala Ile His Lys Arg
        595                 600                 605
```

```
Ser Phe Leu Leu Asn Leu Pro Gly Glu Val Val Thr Lys Asn Asn Lys
    610                 615                 620

Gln Gln Arg Gln Glu Arg Arg Lys Lys Arg Gln Phe Val Arg Ser Gln
625                 630                 635                 640

Ile Arg Met Leu Ala Asn Val Leu Arg Leu Glu Thr Lys Lys Thr Pro
                645                 650                 655

Asp Glu Arg Lys Lys Ala Ile His Lys Leu Met Glu Ile Val Gln Ser
            660                 665                 670

Tyr Asp Ser Trp Thr Ala Ser Gln Lys Glu Val Trp Glu Lys Glu Leu
        675                 680                 685

Asn Leu Leu Thr Asn Met Ala Ala Phe Asn Asp Glu Ile Trp Lys Glu
    690                 695                 700

Ser Leu Val Glu Leu His His Arg Ile Glu Pro Tyr Val Gly Gln Ile
705                 710                 715                 720

Val Ser Lys Trp Arg Lys Gly Leu Ser Glu Gly Arg Lys Asn Leu Ala
                725                 730                 735

Gly Ile Ser Met Trp Asn Ile Asp Glu Leu Glu Asp Thr Arg Arg Leu
            740                 745                 750

Leu Ile Ser Trp Ser Lys Arg Ser Arg Thr Pro Gly Glu Ala Asn Arg
        755                 760                 765

Ile Glu Thr Asp Glu Pro Phe Gly Ser Ser Leu Leu Gln His Ile Gln
    770                 775                 780

Asn Val Lys Asp Asp Arg Leu Lys Gln Met Ala Asn Leu Ile Ile Met
785                 790                 795                 800

Thr Ala Leu Gly Phe Lys Tyr Asp Lys Glu Glu Lys Asp Arg Tyr Lys
                805                 810                 815
```

```
Arg Trp Lys Glu Thr Tyr Pro Ala Cys Gln Ile Ile Leu Phe Glu Asn
            820                 825                 830

Leu Asn Arg Tyr Leu Phe Asn Leu Asp Arg Ser Arg Arg Glu Asn Ser
        835                 840                 845

Arg Leu Met Lys Trp Ala His Arg Ser Ile Pro Arg Thr Val Ser Met
    850                 855                 860

Gln Gly Glu Met Phe Gly Leu Gln Val Gly Asp Val Arg Ser Glu Tyr
865                 870                 875                 880

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

Ala Leu Thr Glu Glu Asp Leu Lys Ala Gly Ser Asn Thr Leu Lys Arg
            900                 905                 910

Leu Ile Glu Asp Gly Phe Ile Asn Glu Ser Glu Leu Ala Tyr Leu Lys
        915                 920                 925

Lys Gly Asp Ile Ile Pro Ser Gln Gly Gly Glu Leu Phe Val Thr Leu
    930                 935                 940

Ser Lys Arg Tyr Lys Lys Asp Ser Asp Asn Asn Glu Leu Thr Val Ile
945                 950                 955                 960

His Ala Asp Ile Asn Ala Ala Gln Asn Leu Gln Lys Arg Phe Trp Gln
                965                 970                 975

Gln Asn Ser Glu Val Tyr Arg Val Pro Cys Gln Leu Ala Arg Met Gly
            980                 985                 990

Glu Asp Lys Leu Tyr Ile Pro Lys  Ser Gln Thr Glu Thr  Ile Lys Lys
        995                 1000                1005

Tyr Phe  Gly Lys Gly Ser Phe  Val Lys Asn Asn Thr  Glu Gln Glu
    1010                1015                1020

Val Tyr  Lys Trp Glu Lys Ser  Glu Lys Met Lys Ile  Lys Thr Asp
    1025                1030                1035
```

Thr Thr Phe Asp Leu Gln Asp Leu Asp Gly Phe Glu Asp Ile Ser
    1040   1045   1050

Lys Thr Ile Glu Leu Ala Gln Glu Gln Gln Lys Lys Tyr Leu Thr
    1055   1060   1065

Met Phe Arg Asp Pro Ser Gly Tyr Phe Phe Asn Asn Glu Thr Trp
    1070   1075   1080

Arg Pro Gln Lys Glu Tyr Trp Ser Ile Val Asn Asn Ile Ile Lys
    1085   1090   1095

Ser Cys Leu Lys Lys Lys Ile Leu Ser Asn Lys Val Glu Leu
    1100   1105   1110

<210> 258
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 5'UTR oligonucleotide
       sequence

<400> 258
gggaaataag agagaaaaga agagtaagaa gaaatataag agccacc   47

<210> 259
<211> 101
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 3'UTR (TriLink standard UTR)
       polynucleotide sequence

<400> 259
gctggagcct cggtggccat gcttcttgcc ccttgggcct ccccccagcc cctcctcccc   60

ttcctgcacc cgtacccccg tggtctttga ataaagtctg a   101

<210> 260
<211> 3420
<212> DNA
<213> Bacillus hisashii

<220>
<221> misc_feature
<222> (1)..(3420)
<223> bhCas12b (V4) polynucleotide sequence

<400> 260
atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc cgccaccaga 60

tccttcatcc tgaagatcga gcccaacgag gaagtgaaga aaggcctctg gaaaacccac 120

gaggtgctga accacggaat cgcctactac atgaatatcc tgaagctgat ccggcaagag 180

gccatctacg agcaccacga gcaggacccc aagaatccca agaaggtgtc caaggccgag 240

atccaggccg agctgtggga tttcgtgctg aagatgcaga agtgcaacag cttcacacac 300

gaggtggaca aggacgaggt gttcaacatc ctgagagagc tgtacgagga actggtgccc 360

agcagcgtgg aaaagaaggg cgaagccaac cagctgagca acaagtttct gtaccctctg 420

gtggacccca acagccagtc tggaaaggga acagccagca gcggcagaaa gcccagatgg 480

tacaacctga agattgccgg cgatccctcc tgggaagaag agaagaagaa gtgggaagaa 540

gataagaaaa aggacccgct ggccaagatc ctgggcaagc tggctgagta cggactgatc 600

cctctgttca tcccctacac cgacagcaac gagcccatcg tgaaagaaat caagtggatg 660

gaaaagtccc ggaaccagag cgtgcggcgg ctggataagg acatgttcat tcaggccctg 720

gaacggttcc tgagctggga gagctggaac ctgaaagtga aagaggaata cgagaaggtc 780

gagaaagagt acaagaccct ggaagagagg atcaaagagg acatccaggc tctgaaggct 840

ctggaacagt atgagaaaga gcggcaagaa cagctgctgc gggacaccct gaacaccaac 900

gagtaccggc tgagcaagag aggccttaga ggctggcggg aaatcatcca gaaatggctg 960

aaaatggacg agaacgagcc ctccgagaag tacctggaag tgttcaagga ctaccagcgg 1020

aagcacccta gagaggccgg cgattacagc gtgtacgagt tcctgtccaa gaaagagaac 1080

cacttcatct ggcggaatca ccctgagtac ccctacctgt acgccacctt ctgcgagatc 1140

gacaagaaaa agaaggacgc caagcagcag gccaccttca cactggccga tcctatcaat 1200

caccctctgt gggtccgatt cgaggaaaga agcggcagca acctgaacaa gtacagaatc 1260

ctgaccgagc agctgcacac cgagaagctg aagaaaaagc tgacagtgca gctggaccgg 1320

ctgatctacc ctacagaatc tggcggctgg gaagagaagg gcaaagtgga cattgtgctg 1380

ctgcccagcc ggcagttcta caaccagatc ttcctggaca tcgaggaaaa gggcaagcac 1440

gccttcacct acaaggatga gagcatcaag ttccctctga agggcacact cggcggagcc 1500

agagtgcagt tcgacagaga tcacctgaga agataccctc acaaggtgga aagcggcaac 1560

gtgggcagaa tctacttcaa catgaccgtg aacatcgagc ctacagagtc cccagtgtcc 1620

aagtctctga agatccaccg ggacgacttc cccaaggtgg tcaacttcaa gcccaaagaa 1680

ctgaccgagt ggatcaagga cagcaagggc aagaaactga agtccggcat cgagtccctg 1740

gaaatcggcc tgagagtgat gagcatcgac ctgggacaga gacaggccgc tgccgcctct 1800

attttcgagg tggtggatca gaagcccgac atcgaaggca agctgttttt cccaatcaag 1860

ggcaccgagc tgtatgccgt gcacagagcc agcttcaaca tcaagctgcc cggcgagaca 1920

ctggtcaaga gcagagaagt gctgcggaag gccagagagg acaatctgaa actgatgaac 1980

cagaagctca acttcctgcg gaacgtgctg cacttccagc agttcgagga catcaccgag 2040

agagagaagc gggtcaccaa gtggatcagc agacaagaga acagcgacgt gcccctggtg 2100

taccaggatg agctgatcca gatccgcgag ctgatgtaca agccttacaa ggactgggtc 2160

gccttcctga agcagctcca caagagactg gaagtcgaga tcggcaaaga agtgaagcac 2220

tggcggaagt ccctgagcga cggaagaaag ggcctgtacg gcatctccct gaagaacatc 2280

gacgagatcg atcggacccg gaagttcctg ctgagatggt ccctgaggcc taccgaacct 2340

ggcgaagtgc gtagactgga acccggccag agattcgcca tcgaccagct gaatcacctg 2400

aacgccctga aagaagatcg gctgaagaag atggccaaca ccatcatcat gcacgccctg 2460

ggctactgct acgacgtgcg gaagaagaaa tggcaggcta agaaccccgc ctgccagatc 2520

atcctgttcg aggatctgag caactacaac ccctacgagg aaaggtcccg cttcgagaac 2580

agcaagctca tgaagtggtc cagacgcgag atccccagac aggttgcact gcagggcgag 2640

atctatggcc tgcaagtggg agaagtgggc gctcagttca gcagcagatt ccacgccaag 2700

acaggcagcc ctggcatcag atgtagcgtc gtgaccaaag agaagctgca ggacaatcgg 2760

ttcttcaaga atctgcagag agagggcaga ctgaccctgg acaaaatcgc cgtgctgaaa 2820

gagggcgatc tgtacccaga caaaggcggc gagaagttca tcagcctgag caaggatcgg 2880

aagtgcgtga ccacacacgc cgacatcaac gccgctcaga acctgcagaa gcggttctgg 2940

acaagaaccc acggcttcta caaggtgtac tgcaaggcct accaggtgga cggccagacc 3000

gtgtacatcc ctgagagcaa ggaccagaag cagaagatca tcgaagagtt cggcgagggc 3060

tacttcattc tgaaggacgg ggtgtacgaa tgggtcaacg ccggcaagct gaaaatcaag 3120

aagggcagct ccaagcagag cagcagcgag ctggtggata gcgacatcct gaaagacagc 3180

ttcgacctgg cctccgagct gaaaggcgaa aagctgatgc tgtacaggga ccccagcggc 3240

aatgtgttcc ccagcgacaa atggatggcc gctggcgtgt tcttcggaaa gctggaacgc 3300

atcctgatca gcaagctgac caaccagtac tccatcagca ccatcgagga cgacagcagc 3360

aagcagtcta tgaaaaggcc ggcggccacg aaaaaggccg gccaggcaaa aaagaaaaag 3420

<210> 261
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 261

Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1 5 10 15

Ala

<210> 262
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 262
atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc c 51

<210> 263
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 101 Cas9 TadAins 1015 polypeptide sequence

<400> 263

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
```

```
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
```

```
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
```

```
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
```

```
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Gly Ser Ser Gly Ser  Glu Thr Pro
    1010                1015                1020

Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser Ser Gly  Ser Glu Val
    1025                1030                1035

Glu Phe  Ser His Glu Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala
    1040                1045                1050

Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro Val Gly  Ala Val Leu
    1055                1060                1065
```

```
Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile
    1070                1075                1080

Gly Leu  His Asp Pro Thr Ala  His Ala Glu Ile Met  Ala Leu Arg
    1085                1090                1095

Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg Leu Ile  Asp Ala Thr
    1100                1105                1110

Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met Cys Ala  Gly Ala Met
    1115                1120                1125

Ile His  Ser Arg Ile Gly Arg  Val Val Phe Gly Val  Arg Asn Ala
    1130                1135                1140

Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp Val Leu  His Tyr Pro
    1145                1150                1155

Gly Met  Asn His Arg Val Glu  Ile Thr Glu Gly Ile  Leu Ala Asp
    1160                1165                1170

Glu Cys  Ala Ala Leu Leu Cys  Tyr Phe Phe Arg Met  Pro Arg Gln
    1175                1180                1185

Val Phe  Asn Ala Gln Lys Lys  Ala Gln Ser Ser Thr  Asp Tyr Asp
    1190                1195                1200

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
```

```
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470
```

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
    1475   1480   1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
    1490   1495   1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
    1505   1510   1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
    1520   1525   1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 264
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 102 Cas9 TadAins 1022
       polypeptide sequence

<400> 264

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

```
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
```

```
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
```

```
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720
```

```
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
```

```
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Gly
    1010                1015                1020


Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser Glu Ser  Ala Thr Pro
    1025                1030                1035


Glu Ser  Ser Gly Ser Glu Val  Glu Phe Ser His Glu  Tyr Trp Met
    1040                1045                1050


Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu
    1055                1060                1065


Val Pro  Val Gly Ala Val Leu  Val Leu Asn Asn Arg  Val Ile Gly
    1070                1075                1080


Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His Asp Pro  Thr Ala His
    1085                1090                1095


Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn
    1100                1105                1110


Tyr Arg  Leu Ile Asp Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys
    1115                1120                1125


Val Met  Cys Ala Gly Ala Met  Ile His Ser Arg Ile  Gly Arg Val
    1130                1135                1140
```

```
Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu
    1145                1150                1155

Met Asp  Val Leu His Tyr Pro  Gly Met Asn His Arg  Val Glu Ile
    1160                1165                1170

Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr
    1175                1180                1185

Phe Phe  Arg Met Pro Arg Gln  Val Phe Asn Ala Gln  Lys Lys Ala
    1190                1195                1200

Gln Ser  Ser Thr Asp Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350
```

```
Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545
```

Ser Gln Leu Gly Gly Asp
    1550

<210> 265
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 103 Cas9 TadAins 1029
       polypeptide sequence

<400> 265

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145   150   155   160

```
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
```

```
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
```

```
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
```

```
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020
```

```
Lys Ser  Glu Gln Glu Ile Gly  Ser Ser Gly Ser Glu  Thr Pro Gly
    1025                1030                1035

Thr Ser  Glu Ser Ala Thr Pro  Glu Ser Ser Gly Ser  Glu Val Glu
    1040                1045                1050

Phe Ser  His Glu Tyr Trp Met  Arg His Ala Leu Thr  Leu Ala Lys
    1055                1060                1065

Arg Ala  Arg Asp Glu Arg Glu  Val Pro Val Gly Ala  Val Leu Val
    1070                1075                1080

Leu Asn  Asn Arg Val Ile Gly  Glu Gly Trp Asn Arg  Ala Ile Gly
    1085                1090                1095

Leu His  Asp Pro Thr Ala His  Ala Glu Ile Met Ala  Leu Arg Gln
    1100                1105                1110

Gly Gly  Leu Val Met Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu
    1115                1120                1125

Tyr Val  Thr Phe Glu Pro Cys  Val Met Cys Ala Gly  Ala Met Ile
    1130                1135                1140

His Ser  Arg Ile Gly Arg Val  Val Phe Gly Val Arg  Asn Ala Lys
    1145                1150                1155

Thr Gly  Ala Ala Gly Ser Leu  Met Asp Val Leu His  Tyr Pro Gly
    1160                1165                1170

Met Asn  His Arg Val Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu
    1175                1180                1185

Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val
    1190                1195                1200

Phe Asn  Ala Gln Lys Lys Ala  Gln Ser Ser Thr Asp  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
```

```
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425
```

```
Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440


Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455


Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470


Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485


Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500


Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515


Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530


Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545


Ser Gln  Leu Gly Gly Asp
    1550


<210>  266
<211>  1554
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 103 Cas9 TadAins 1040
       polypeptide sequence

<400>  266

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30
```

```
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
```

```
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460
```

```
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670
```

```
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
```

```
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Gly Ser Ser Gly Ser  Glu Thr Pro Gly Thr  Ser Glu Ser
    1040                1045                1050

Ala Thr  Pro Glu Ser Ser Gly  Ser Glu Val Glu Phe  Ser His Glu
    1055                1060                1065

Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp
    1070                1075                1080

Glu Arg  Glu Val Pro Val Gly  Ala Val Leu Val Leu  Asn Asn Arg
    1085                1090                1095
```

```
Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro
    1100                1105                1110

Thr Ala  His Ala Glu Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val
    1115                1120                1125

Met Gln  Asn Tyr Arg Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe
    1130                1135                1140

Glu Pro  Cys Val Met Cys Ala  Gly Ala Met Ile His  Ser Arg Ile
    1145                1150                1155

Gly Arg  Val Val Phe Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala
    1160                1165                1170

Gly Ser  Leu Met Asp Val Leu  His Tyr Pro Gly Met  Asn His Arg
    1175                1180                1185

Val Glu  Ile Thr Glu Gly Ile  Leu Ala Asp Glu Cys  Ala Ala Leu
    1190                1195                1200

Leu Cys  Tyr Phe Phe Arg Met  Pro Arg Gln Val Phe  Asn Ala Gln
    1205                1210                1215

Lys Lys  Ala Gln Ser Ser Thr  Asp Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305
```

```
Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500
```

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
    1505   1510   1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
    1520   1525   1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 267
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 105 Cas9 TadAins 1068
       polypeptide sequence

<400> 267

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

```
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
```

```
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540
```

```
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560


Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575


Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590


Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605


Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620


Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640


His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655


Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670


Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685


Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700


Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720


His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735


Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750


Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
```

```
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
```

```
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Gly Ser  Ser Gly Ser Glu Thr  Pro Gly Thr Ser Glu  Ser Ala Thr
    1070                1075                1080

Pro Glu  Ser Ser Gly Ser Glu  Val Glu Phe Ser His  Glu Tyr Trp
    1085                1090                1095

Met Arg  His Ala Leu Thr Leu  Ala Lys Arg Ala Arg  Asp Glu Arg
    1100                1105                1110

Glu Val  Pro Val Gly Ala Val  Leu Val Leu Asn Asn  Arg Val Ile
    1115                1120                1125

Gly Glu  Gly Trp Asn Arg Ala  Ile Gly Leu His Asp  Pro Thr Ala
    1130                1135                1140

His Ala  Glu Ile Met Ala Leu  Arg Gln Gly Gly Leu  Val Met Gln
    1145                1150                1155

Asn Tyr  Arg Leu Ile Asp Ala  Thr Leu Tyr Val Thr  Phe Glu Pro
    1160                1165                1170

Cys Val  Met Cys Ala Gly Ala  Met Ile His Ser Arg  Ile Gly Arg
```

```
    1175                1180                1185

Val Val  Phe Gly Val Arg Asn  Ala Lys Thr Gly Ala  Ala Gly Ser
    1190                1195                1200

Leu Met  Asp Val Leu His Tyr  Pro Gly Met Asn His  Arg Val Glu
    1205                1210                1215

Ile Thr  Glu Gly Ile Leu Ala  Asp Glu Cys Ala Ala  Leu Leu Cys
    1220                1225                1230

Tyr Phe  Phe Arg Met Pro Arg  Gln Val Phe Asn Ala  Gln Lys Lys
    1235                1240                1245

Ala Gln  Ser Ser Thr Asp Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380
```

```
Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545

Ser Gln  Leu Gly Gly Asp
    1550
```

<210> 268
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 106 Cas9 TadAins 1247 polypeptide sequence

<400> 268

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60


Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80


Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95


Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110


His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125


His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140


Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160


Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175


Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190
```

```
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
```

```
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620
```

```
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
```

```
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050
```

```
Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Gly
    1235                1240                1245

Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser Glu Ser  Ala Thr Pro
    1250                1255                1260
```

```
Glu Ser  Ser Gly Ser Glu Val  Glu Phe Ser His Glu  Tyr Trp Met
    1265                1270                1275

Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu
    1280                1285                1290

Val Pro  Val Gly Ala Val Leu  Val Leu Asn Asn Arg  Val Ile Gly
    1295                1300                1305

Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His Asp Pro  Thr Ala His
    1310                1315                1320

Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn
    1325                1330                1335

Tyr Arg  Leu Ile Asp Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys
    1340                1345                1350

Val Met  Cys Ala Gly Ala Met  Ile His Ser Arg Ile  Gly Arg Val
    1355                1360                1365

Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu
    1370                1375                1380

Met Asp  Val Leu His Tyr Pro  Gly Met Asn His Arg  Val Glu Ile
    1385                1390                1395

Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr
    1400                1405                1410

Phe Phe  Arg Met Pro Arg Gln  Val Phe Asn Ala Gln  Lys Lys Ala
    1415                1420                1425

Gln Ser  Ser Thr Asp Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455
```

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
    1460   1465   1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
    1475   1480   1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
    1490   1495   1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
    1505   1510   1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
    1520   1525   1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 269
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 107 Cas9 TadAins 1054
       polypeptide sequence

<400> 269

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

```
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
```

```
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
```

```
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
```

```
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925
```

```
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Ser Ser Gly Ser Glu  Thr Pro Gly Thr Ser  Glu Ser Ala
    1055                1060                1065

Thr Pro  Glu Ser Ser Gly Ser  Glu Val Glu Phe Ser  His Glu Tyr
    1070                1075                1080

Trp Met  Arg His Ala Leu Thr  Leu Ala Lys Arg Ala  Arg Asp Glu
    1085                1090                1095

Arg Glu  Val Pro Val Gly Ala  Val Leu Val Leu Asn  Asn Arg Val
    1100                1105                1110

Ile Gly  Glu Gly Trp Asn Arg  Ala Ile Gly Leu His  Asp Pro Thr
    1115                1120                1125

Ala His  Ala Glu Ile Met Ala  Leu Arg Gln Gly Gly  Leu Val Met
```

```
    1130                1135                1140

Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu Tyr Val  Thr Phe Glu
    1145                1150                1155

Pro Cys  Val Met Cys Ala Gly  Ala Met Ile His Ser  Arg Ile Gly
    1160                1165                1170

Arg Val  Val Phe Gly Val Arg  Asn Ala Lys Thr Gly  Ala Ala Gly
    1175                1180                1185

Ser Leu  Met Asp Val Leu His  Tyr Pro Gly Met Asn  His Arg Val
    1190                1195                1200

Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu Cys Ala  Ala Leu Leu
    1205                1210                1215

Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val Phe Asn  Ala Gln Lys
    1220                1225                1230

Lys Ala  Gln Ser Ser Thr Asp  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335
```

```
Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
```

1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 270
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 108 Cas9 TadAins 1026 polypeptide sequence

<400> 270

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

```
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575
```

```
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
```

```
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005
```

```
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gly Ser Ser Gly  Ser Glu Thr Pro Gly  Thr Ser Glu
    1025                1030                1035

Ser Ala  Thr Pro Glu Ser Ser  Gly Ser Glu Val Glu  Phe Ser His
    1040                1045                1050

Glu Tyr  Trp Met Arg His Ala  Leu Thr Leu Ala Lys  Arg Ala Arg
    1055                1060                1065

Asp Glu  Arg Glu Val Pro Val  Gly Ala Val Leu Val  Leu Asn Asn
    1070                1075                1080

Arg Val  Ile Gly Glu Gly Trp  Asn Arg Ala Ile Gly  Leu His Asp
    1085                1090                1095

Pro Thr  Ala His Ala Glu Ile  Met Ala Leu Arg Gln  Gly Gly Leu
    1100                1105                1110

Val Met  Gln Asn Tyr Arg Leu  Ile Asp Ala Thr Leu  Tyr Val Thr
    1115                1120                1125

Phe Glu  Pro Cys Val Met Cys  Ala Gly Ala Met Ile  His Ser Arg
    1130                1135                1140

Ile Gly  Arg Val Val Phe Gly  Val Arg Asn Ala Lys  Thr Gly Ala
    1145                1150                1155

Ala Gly  Ser Leu Met Asp Val  Leu His Tyr Pro Gly  Met Asn His
    1160                1165                1170

Arg Val  Glu Ile Thr Glu Gly  Ile Leu Ala Asp Glu  Cys Ala Ala
    1175                1180                1185

Leu Leu  Cys Tyr Phe Phe Arg  Met Pro Arg Gln Val  Phe Asn Ala
    1190                1195                1200

Gln Lys  Lys Ala Gln Ser Ser  Thr Asp Gln Glu Ile  Gly Lys Ala
    1205                1210                1215
```

```
Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410
```

```
Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425


Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440


Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455


Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470


Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485


Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500


Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515


Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530


Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545


Ser Gln  Leu Gly Gly Asp
    1550


<210>  271
<211>  1540
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 109 Cas9 TadAins 768
       polypeptide sequence

<400>  271

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15
```

```
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
```

```
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445
```

```
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
```

```
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
    770                 775                 780

Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
785                 790                 795                 800

His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro
                805                 810                 815

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp
            820                 825                 830

Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met
        835                 840                 845

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
    850                 855                 860

Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala
865                 870                 875                 880
```

```
Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala
               885                 890                 895


Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly
            900                 905                 910


Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
        915                 920                 925


Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Thr Thr Gln Lys
    930                 935                 940


Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile
945                 950                 955                 960


Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr
               965                 970                 975


Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg
            980                 985                 990


Asp Met Tyr Val Asp Gln Glu Leu  Asp Ile Asn Arg Leu  Ser Asp Tyr
        995                 1000                1005


Asp Val  Asp His Ile Val Pro  Gln Ser Phe Leu Lys  Asp Asp Ser
    1010                1015                1020


Ile Asp  Asn Lys Val Leu Thr  Arg Ser Asp Lys Asn  Arg Gly Lys
    1025                1030                1035


Ser Asp  Asn Val Pro Ser Glu  Glu Val Val Lys Lys  Met Lys Asn
    1040                1045                1050


Tyr Trp  Arg Gln Leu Leu Asn  Ala Lys Leu Ile Thr  Gln Arg Lys
    1055                1060                1065


Phe Asp  Asn Leu Thr Lys Ala  Glu Arg Gly Gly Leu  Ser Glu Leu
    1070                1075                1080


Asp Lys  Ala Gly Phe Ile Lys  Arg Gln Leu Val Glu  Thr Arg Gln
```

```
    1085                1090                1095

Ile Thr  Lys His Val Ala Gln  Ile Leu Asp Ser Arg  Met Asn Thr
    1100                1105                1110

Lys Tyr  Asp Glu Asn Asp Lys  Leu Ile Arg Glu Val  Lys Val Ile
    1115                1120                1125

Thr Leu  Lys Ser Lys Leu Val  Ser Asp Phe Arg Lys  Asp Phe Gln
    1130                1135                1140

Phe Tyr  Lys Val Arg Glu Ile  Asn Asn Tyr His His  Ala His Asp
    1145                1150                1155

Ala Tyr  Leu Asn Ala Val Val  Gly Thr Ala Leu Ile  Lys Lys Tyr
    1160                1165                1170

Pro Lys  Leu Glu Ser Glu Phe  Val Tyr Gly Asp Tyr  Lys Val Tyr
    1175                1180                1185

Asp Val  Arg Lys Met Ile Ala  Lys Ser Glu Gln Glu  Ile Gly Lys
    1190                1195                1200

Ala Thr  Ala Lys Tyr Phe Phe  Tyr Ser Asn Ile Met  Asn Phe Phe
    1205                1210                1215

Lys Thr  Glu Ile Thr Leu Ala  Asn Gly Glu Ile Arg  Lys Arg Pro
    1220                1225                1230

Leu Ile  Glu Thr Asn Gly Glu  Thr Gly Glu Ile Val  Trp Asp Lys
    1235                1240                1245

Gly Arg  Asp Phe Ala Thr Val  Arg Lys Val Leu Ser  Met Pro Gln
    1250                1255                1260

Val Asn  Ile Val Lys Lys Thr  Glu Val Gln Thr Gly  Gly Phe Ser
    1265                1270                1275

Lys Glu  Ser Ile Leu Pro Lys  Arg Asn Ser Asp Lys  Leu Ile Ala
    1280                1285                1290
```

```
Arg Lys  Lys Asp Trp Asp Pro  Lys Lys Tyr Gly Gly  Phe Asp Ser
    1295                1300                1305

Pro Thr  Val Ala Tyr Ser Val  Leu Val Val Ala Lys  Val Glu Lys
    1310                1315                1320

Gly Lys  Ser Lys Lys Leu Lys  Ser Val Lys Glu Leu  Leu Gly Ile
    1325                1330                1335

Thr Ile  Met Glu Arg Ser Ser  Phe Glu Lys Asn Pro  Ile Asp Phe
    1340                1345                1350

Leu Glu  Ala Lys Gly Tyr Lys  Glu Val Lys Lys Asp  Leu Ile Ile
    1355                1360                1365

Lys Leu  Pro Lys Tyr Ser Leu  Phe Glu Leu Glu Asn  Gly Arg Lys
    1370                1375                1380

Arg Met  Leu Ala Ser Ala Gly  Glu Leu Gln Lys Gly  Asn Glu Leu
    1385                1390                1395

Ala Leu  Pro Ser Lys Tyr Val  Asn Phe Leu Tyr Leu  Ala Ser His
    1400                1405                1410

Tyr Glu  Lys Leu Lys Gly Ser  Pro Glu Asp Asn Glu  Gln Lys Gln
    1415                1420                1425

Leu Phe  Val Glu Gln His Lys  His Tyr Leu Asp Glu  Ile Ile Glu
    1430                1435                1440

Gln Ile  Ser Glu Phe Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn
    1445                1450                1455

Leu Asp  Lys Val Leu Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro
    1460                1465                1470

Ile Arg  Glu Gln Ala Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr
    1475                1480                1485

Asn Leu  Gly Ala Pro Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile
```

1490 1495 1500

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr
1505 1510 1515

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1520 1525 1530

Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 272
<211> 1540
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 110.1 Cas9 TadAins 1250 polypeptide sequence

<400> 272

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

```
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335
```

```
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540
```

```
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560


Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575


Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590


Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605


Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620


Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640


His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655


Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670


Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685


Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700


Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720


His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735


Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750


Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765
```

```
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
```

```
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080


Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095


Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110


Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125


Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140


Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155


Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170


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

```
Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Gly  Ser Ser Gly Ser Glu  Thr Pro Gly Thr Ser  Glu Ser Ala
    1250                1255                1260

Thr Pro  Glu Ser Ser Gly Ser  Glu Val Glu Phe Ser  His Glu Tyr
    1265                1270                1275

Trp Met  Arg His Ala Leu Thr  Leu Ala Lys Arg Ala  Arg Asp Glu
    1280                1285                1290

Arg Glu  Val Pro Val Gly Ala  Val Leu Val Leu Asn  Asn Arg Val
    1295                1300                1305

Ile Gly  Glu Gly Trp Asn Arg  Ala Ile Gly Leu His  Asp Pro Thr
    1310                1315                1320

Ala His  Ala Glu Ile Met Ala  Leu Arg Gln Gly Gly  Leu Val Met
    1325                1330                1335

Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu Tyr Val  Thr Phe Glu
    1340                1345                1350

Pro Cys  Val Met Cys Ala Gly  Ala Met Ile His Ser  Arg Ile Gly
    1355                1360                1365

Arg Val  Val Phe Gly Val Arg  Asn Ala Lys Thr Gly  Ala Ala Gly
    1370                1375                1380
```

```
Ser Leu  Met Asp Val Leu His  Tyr Pro Gly Met Asn  His Arg Val
    1385                1390                1395


Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu Cys Ala  Ala Leu Leu
    1400                1405                1410


Cys Tyr  Phe Phe Arg Met Pro  Arg Glu Asp Asn Glu  Gln Lys Gln
    1415                1420                1425


Leu Phe  Val Glu Gln His Lys  His Tyr Leu Asp Glu  Ile Ile Glu
    1430                1435                1440


Gln Ile  Ser Glu Phe Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn
    1445                1450                1455


Leu Asp  Lys Val Leu Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro
    1460                1465                1470


Ile Arg  Glu Gln Ala Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr
    1475                1480                1485


Asn Leu  Gly Ala Pro Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile
    1490                1495                1500


Asp Arg  Lys Arg Tyr Thr Ser  Thr Lys Glu Val Leu  Asp Ala Thr
    1505                1510                1515


Leu Ile  His Gln Ser Ile Thr  Gly Leu Tyr Glu Thr  Arg Ile Asp
    1520                1525                1530


Leu Ser  Gln Leu Gly Gly Asp
    1535                1540


<210>  273
<211>  1543
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 110.2 Cas9 TadAins 1250
       polypeptide sequence

<400>  273
```

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
```

```
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430
```

```
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
```

```
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860
```

```
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880


Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895


Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910


Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925


Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940


Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
```

```
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Gly  Ser Ser Gly Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser
    1250                1255                1260

Glu Ser  Ala Thr Pro Glu Ser  Ser Gly Ser Glu Val  Glu Phe Ser
    1265                1270                1275
```

```
His Glu  Tyr Trp Met Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala
    1280                1285                1290

Arg Asp  Glu Arg Glu Val Pro  Val Gly Ala Val Leu  Val Leu Asn
    1295                1300                1305

Asn Arg  Val Ile Gly Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His
    1310                1315                1320

Asp Pro  Thr Ala His Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly
    1325                1330                1335

Leu Val  Met Gln Asn Tyr Arg  Leu Ile Asp Ala Thr  Leu Tyr Val
    1340                1345                1350

Thr Phe  Glu Pro Cys Val Met  Cys Ala Gly Ala Met  Ile His Ser
    1355                1360                1365

Arg Ile  Gly Arg Val Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly
    1370                1375                1380

Ala Ala  Gly Ser Leu Met Asp  Val Leu His Tyr Pro  Gly Met Asn
    1385                1390                1395

His Arg  Val Glu Ile Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala
    1400                1405                1410

Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro Arg Glu  Asp Asn Glu
    1415                1420                1425

Gln Lys  Gln Leu Phe Val Glu  Gln His Lys His Tyr  Leu Asp Glu
    1430                1435                1440

Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys Arg Val  Ile Leu Ala
    1445                1450                1455

Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala Tyr Asn  Lys His Arg
    1460                1465                1470

Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn Ile Ile  His Leu Phe
```

1475 1480 1485

Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
1490 1495 1500

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu
1505 1510 1515

Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr
1520 1525 1530

Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 274
<211> 1546
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 110.3 Cas9 TadAins 1250 polypeptide sequence

<400> 274

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

```
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
```

```
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
```

```
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
```

```
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
```

```
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
               965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
           980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170
```

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Gly  Ser Ser Gly Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser
    1250                1255                1260

Glu Ser  Ala Thr Pro Glu Ser  Gly Ser Ser Ser Gly  Ser Glu Val
    1265                1270                1275

Glu Phe  Ser His Glu Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala
    1280                1285                1290

Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro Val Gly  Ala Val Leu
    1295                1300                1305

Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile
    1310                1315                1320

Gly Leu  His Asp Pro Thr Ala  His Ala Glu Ile Met  Ala Leu Arg
    1325                1330                1335

Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg Leu Ile  Asp Ala Thr
    1340                1345                1350

Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met Cys Ala  Gly Ala Met
    1355                1360                1365
```

```
Ile His  Ser Arg Ile Gly Arg  Val Val Phe Gly Val  Arg Asn Ala
    1370                1375                1380

Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp Val Leu  His Tyr Pro
    1385                1390                1395

Gly Met  Asn His Arg Val Glu  Ile Thr Glu Gly Ile  Leu Ala Asp
    1400                1405                1410

Glu Cys  Ala Ala Leu Leu Cys  Tyr Phe Phe Arg Met  Pro Arg Glu
    1415                1420                1425

Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu Gln His  Lys His Tyr
    1430                1435                1440

Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val
    1445                1450                1455

Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn
    1460                1465                1470

Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile
    1475                1480                1485

His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys
    1490                1495                1500

Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys
    1505                1510                1515

Glu Val  Leu Asp Ala Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu
    1520                1525                1530

Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540                1545
```

<210> 275
<211> 1546
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of artificial sequence: 110.4 Cas9 TadAins 1250 polypeptide sequence

<400> 275

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60


Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80


Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95


Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110


His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125


His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140


Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160


Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175


Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190


Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
```

```
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
```

```
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
```

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

```
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
```

```
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Gly  Ser Ser Gly Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser
    1250                1255                1260
```

```
Glu Ser  Ala Thr Pro Glu Ser  Gly Ser Ser Ser Gly  Ser Glu Val
    1265                1270                1275

Glu Phe  Ser His Glu Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala
    1280                1285                1290

Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro Val Gly  Ala Val Leu
    1295                1300                1305

Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile
    1310                1315                1320

Gly Leu  His Asp Pro Thr Ala  His Ala Glu Ile Met  Ala Leu Arg
    1325                1330                1335

Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg Leu Ile  Asp Ala Thr
    1340                1345                1350

Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met Cys Ala  Gly Ala Met
    1355                1360                1365

Ile His  Ser Arg Ile Gly Arg  Val Val Phe Gly Val  Arg Asn Ala
    1370                1375                1380

Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp Val Leu  His Tyr Pro
    1385                1390                1395

Gly Met  Asn His Arg Val Glu  Ile Thr Glu Gly Ile  Leu Ala Asp
    1400                1405                1410

Glu Cys  Ala Ala Leu Leu Cys  Tyr Phe Phe Arg Met  Arg Arg Glu
    1415                1420                1425

Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu Gln His  Lys His Tyr
    1430                1435                1440

Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val
    1445                1450                1455

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

```
    1460                1465                1470

Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile
    1475                1480                1485

His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys
    1490                1495                1500

Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys
    1505                1510                1515

Glu Val  Leu Asp Ala Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu
    1520                1525                1530

Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540                1545
```

<210> 276
<211> 1546
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 110.5 Cas9 TadAins 1249 polypeptide sequence

<400> 276

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80
```

```
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300
```

```
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
```

```
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735
```

```
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
```

```
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080


Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095


Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110


Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125


Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140


Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155
```

```
Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Gly Ser  Ser Gly Ser Ser Gly  Ser Glu Thr Pro Gly  Thr Ser Glu
    1250                1255                1260

Ser Ala  Thr Pro Glu Ser Gly  Ser Ser Ser Gly Ser  Glu Val Glu
    1265                1270                1275

Phe Ser  His Glu Tyr Trp Met  Arg His Ala Leu Thr  Leu Ala Lys
    1280                1285                1290

Arg Ala  Arg Asp Glu Arg Glu  Val Pro Val Gly Ala  Val Leu Val
    1295                1300                1305

Leu Asn  Asn Arg Val Ile Gly  Glu Gly Trp Asn Arg  Ala Ile Gly
    1310                1315                1320

Leu His  Asp Pro Thr Ala His  Ala Glu Ile Met Ala  Leu Arg Gln
    1325                1330                1335

Gly Gly  Leu Val Met Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu
    1340                1345                1350
```

```
Tyr Val  Thr Phe Glu Pro Cys  Val Met Cys Ala Gly  Ala Met Ile
    1355                1360                1365

His Ser  Arg Ile Gly Arg Val  Val Phe Gly Val Arg  Asn Ala Lys
    1370                1375                1380

Thr Gly  Ala Ala Gly Ser Leu  Met Asp Val Leu His  Tyr Pro Gly
    1385                1390                1395

Met Asn  His Arg Val Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu
    1400                1405                1410

Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg Met Arg  Arg Pro Glu
    1415                1420                1425

Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu Gln His  Lys His Tyr
    1430                1435                1440

Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val
    1445                1450                1455

Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn
    1460                1465                1470

Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile
    1475                1480                1485

His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys
    1490                1495                1500

Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys
    1505                1510                1515

Glu Val  Leu Asp Ala Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu
    1520                1525                1530

Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540                1545
```

<210> 277
<211> 1551

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 110.5 Cas9 TadAins delta 59-66 1250 polypeptide sequence

<400> 277

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

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

```
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400
```

```
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
```

```
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
```

```
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
```

```
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245
```

```
Pro Gly  Ser Ser Gly Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser
    1250                1255                1260

Glu Ser  Ala Thr Pro Glu Ser  Gly Ser Ser Gly Ser  Glu Val Glu
    1265                1270                1275

Phe Ser  His Glu Tyr Trp Met  Arg His Ala Leu Thr  Leu Ala Lys
    1280                1285                1290

Arg Ala  Arg Asp Glu Arg Glu  Val Pro Val Gly Ala  Val Leu Val
    1295                1300                1305

Leu Asn  Asn Arg Val Ile Gly  Glu Gly Trp Asn Arg  Ala His Ala
    1310                1315                1320

Glu Ile  Met Ala Leu Arg Gln  Gly Gly Leu Val Met  Gln Asn Tyr
    1325                1330                1335

Arg Leu  Ile Asp Ala Thr Leu  Tyr Val Thr Phe Glu  Pro Cys Val
    1340                1345                1350

Met Cys  Ala Gly Ala Met Ile  His Ser Arg Ile Gly  Arg Val Val
    1355                1360                1365

Phe Gly  Val Arg Asn Ala Lys  Thr Gly Ala Ala Gly  Ser Leu Met
    1370                1375                1380

Asp Val  Leu His Tyr Pro Gly  Met Asn His Arg Val  Glu Ile Thr
    1385                1390                1395

Glu Gly  Ile Leu Ala Asp Glu  Cys Ala Ala Leu Leu  Cys Tyr Phe
    1400                1405                1410

Phe Arg  Met Pro Arg Gln Val  Phe Asn Ala Gln Lys  Lys Ala Gln
    1415                1420                1425

Ser Ser  Thr Asp Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1430                1435                1440

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
```

```
    1445                1450                1455


Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1460                1465                1470


Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1475                1480                1485


Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1490                1495                1500


Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg
    1505                1510                1515


Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1520                1525                1530


Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1535                1540                1545


Gly Gly  Asp
    1550


<210>  278
<211>  1546
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 110.6 Cas9 TadAins 1251
       polypeptide sequence

<400>  278

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
```

```
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
```

```
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480
```

```
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

```
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
```

```
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925


Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940


Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080


Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095


Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110


Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125
```

```
Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Gly Ser Ser Gly Ser  Ser Gly Ser Glu Thr  Pro Gly Thr
    1250                1255                1260

Ser Glu  Ser Ala Thr Pro Glu  Ser Gly Ser Ser Ser  Gly Ser Glu
    1265                1270                1275

Val Glu  Phe Ser His Glu Tyr  Trp Met Arg His Ala  Leu Thr Leu
    1280                1285                1290

Ala Lys  Arg Ala Arg Asp Glu  Arg Glu Val Pro Val  Gly Ala Val
    1295                1300                1305

Leu Val  Leu Asn Asn Arg Val  Ile Gly Glu Gly Trp  Asn Arg Ala
    1310                1315                1320
```

```
Ile Gly  Leu His Asp Pro Thr  Ala His Ala Glu Ile  Met Ala Leu
    1325                 1330                 1335

Arg Gln  Gly Gly Leu Val Met  Gln Asn Tyr Arg Leu  Ile Asp Ala
    1340                 1345                 1350

Thr Leu  Tyr Val Thr Phe Glu  Pro Cys Val Met Cys  Ala Gly Ala
    1355                 1360                 1365

Met Ile  His Ser Arg Ile Gly  Arg Val Val Phe Gly  Val Arg Asn
    1370                 1375                 1380

Ala Lys  Thr Gly Ala Ala Gly  Ser Leu Met Asp Val  Leu His Tyr
    1385                 1390                 1395

Pro Gly  Met Asn His Arg Val  Glu Ile Thr Glu Gly  Ile Leu Ala
    1400                 1405                 1410

Asp Glu  Cys Ala Ala Leu Leu  Cys Tyr Phe Phe Arg  Met Arg Arg
    1415                 1420                 1425

Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu Gln His  Lys His Tyr
    1430                 1435                 1440

Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val
    1445                 1450                 1455

Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn
    1460                 1465                 1470

Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile
    1475                 1480                 1485

His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys
    1490                 1495                 1500

Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys
    1505                 1510                 1515

Glu Val  Leu Asp Ala Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu
    1520                 1525                 1530
```

Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
    1535   1540   1545

<210> 279
<211> 1546
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 110.7 Cas9 TadAins 1252
       polypeptide sequence

<400> 279

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

```
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
```

```
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
```

```
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
```

```
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215
```

```
Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Gly Ser Ser Gly  Ser Ser Gly Ser Glu  Thr Pro Gly
    1250                1255                1260

Thr Ser  Glu Ser Ala Thr Pro  Glu Ser Gly Ser Ser  Ser Gly Ser
    1265                1270                1275

Glu Val  Glu Phe Ser His Glu  Tyr Trp Met Arg His  Ala Leu Thr
    1280                1285                1290

Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu Val Pro  Val Gly Ala
    1295                1300                1305

Val Leu  Val Leu Asn Asn Arg  Val Ile Gly Glu Gly  Trp Asn Arg
    1310                1315                1320

Ala Ile  Gly Leu His Asp Pro  Thr Ala His Ala Glu  Ile Met Ala
    1325                1330                1335

Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn Tyr Arg  Leu Ile Asp
    1340                1345                1350

Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys Val Met  Cys Ala Gly
    1355                1360                1365

Ala Met  Ile His Ser Arg Ile  Gly Arg Val Val Phe  Gly Val Arg
    1370                1375                1380

Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu Met Asp  Val Leu His
    1385                1390                1395

Tyr Pro  Gly Met Asn His Arg  Val Glu Ile Thr Glu  Gly Ile Leu
    1400                1405                1410

Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Arg
```

```
    1415                1420                1425


Arg Asn  Glu Gln Lys Gln Leu  Phe Val Glu Gln His  Lys His Tyr
    1430                1435                1440


Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val
    1445                1450                1455


Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn
    1460                1465                1470


Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile
    1475                1480                1485


His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys
    1490                1495                1500


Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys
    1505                1510                1515


Glu Val  Leu Asp Ala Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu
    1520                1525                1530


Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540                1545


<210>  280
<211>  1533
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 110.8 Cas9 TadAins delta
       59-66 C-truncate 1250 polypeptide sequence

<400>  280

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30
```

```
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255
```

```
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460
```

```
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
```

```
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
```

```
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110
```

```
Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Gly  Ser Ser Gly Ser Glu  Thr Pro Gly Thr Ser  Glu Ser Ala
    1250                1255                1260

Thr Pro  Glu Ser Ser Gly Ser  Glu Val Glu Phe Ser  His Glu Tyr
    1265                1270                1275

Trp Met  Arg His Ala Leu Thr  Leu Ala Lys Arg Ala  Arg Asp Glu
    1280                1285                1290

Arg Glu  Val Pro Val Gly Ala  Val Leu Val Leu Asn  Asn Arg Val
    1295                1300                1305
```

```
Ile Gly  Glu Gly Trp Asn Arg  Ala His Ala Glu Ile  Met Ala Leu
    1310                 1315                 1320

Arg Gln  Gly Gly Leu Val Met  Gln Asn Tyr Arg Leu  Ile Asp Ala
    1325                 1330                 1335

Thr Leu  Tyr Val Thr Phe Glu  Pro Cys Val Met Cys  Ala Gly Ala
    1340                 1345                 1350

Met Ile  His Ser Arg Ile Gly  Arg Val Val Phe Gly  Val Arg Asn
    1355                 1360                 1365

Ala Lys  Thr Gly Ala Ala Gly  Ser Leu Met Asp Val  Leu His Tyr
    1370                 1375                 1380

Pro Gly  Met Asn His Arg Val  Glu Ile Thr Glu Gly  Ile Leu Ala
    1385                 1390                 1395

Asp Glu  Cys Ala Ala Leu Leu  Cys Tyr Phe Phe Arg  Met Pro Arg
    1400                 1405                 1410

Gln Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1415                 1420                 1425

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1430                 1435                 1440

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1445                 1450                 1455

Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1460                 1465                 1470

Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1475                 1480                 1485

Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1490                 1495                 1500

Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1505                 1510                 1515
```

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
    1520   1525   1530

<210> 281
<211> 1549
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 111.1 Cas9 TadAins 997
       polypeptide sequence

<400> 281

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

```
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
```

```
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
```

```
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ser His Glu  Tyr Trp Met Arg His  Ala Leu Thr
        995                 1000                1005

Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu Val Pro  Val Gly Ala
```

```
    1010                1015                1020

Val Leu  Val Leu Asn Asn Arg  Val Ile Gly Glu Gly  Trp Asn Arg
    1025                1030                1035

Ala Ile  Gly Leu His Asp Pro  Thr Ala His Ala Glu  Ile Met Ala
    1040                1045                1050

Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn Tyr Arg  Leu Ile Asp
    1055                1060                1065

Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys Val Met  Cys Ala Gly
    1070                1075                1080

Ala Met  Ile His Ser Arg Ile  Gly Arg Val Val Phe  Gly Val Arg
    1085                1090                1095

Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu Met Asp  Val Leu His
    1100                1105                1110

Tyr Pro  Gly Met Asn His Arg  Val Glu Ile Thr Glu  Gly Ile Leu
    1115                1120                1125

Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro
    1130                1135                1140

Arg Gln  Val Phe Asn Ala Gln  Lys Lys Ala Gln Ser  Ser Thr Asp
    1145                1150                1155

Gly Ser  Ser Gly Ser Glu Thr  Pro Gly Thr Ser Glu  Ser Ala Thr
    1160                1165                1170

Pro Glu  Ser Ser Gly Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu
    1175                1180                1185

Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile
    1190                1195                1200

Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe
    1205                1210                1215
```

```
Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu
    1220                1225                1230

Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly
    1235                1240                1245

Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr
    1250                1255                1260

Val Arg  Lys Val Leu Ser Met  Pro Gln Val Asn Ile  Val Lys Lys
    1265                1270                1275

Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro
    1280                1285                1290

Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp
    1295                1300                1305

Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser
    1310                1315                1320

Val Leu  Val Val Ala Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu
    1325                1330                1335

Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser
    1340                1345                1350

Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr
    1355                1360                1365

Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser
    1370                1375                1380

Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala
    1385                1390                1395

Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr
    1400                1405                1410

Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly
```

```
    1415                1420                1425

Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu Phe Val  Glu Gln His
    1430                1435                1440

Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser
    1445                1450                1455

Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser
    1460                1465                1470

Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu
    1475                1480                1485

Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala
    1490                1495                1500

Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr
    1505                1510                1515

Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu Ile His  Gln Ser Ile
    1520                1525                1530

Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu Ser Gln  Leu Gly Gly
    1535                1540                1545

Asp
```

<210> 282
<211> 1555
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 111.2 Cas9 TadAins 997 polypeptide sequence

<400> 282

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15
```

```
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
```

```
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445
```

```
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670
```

```
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
```

```
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ser His Glu  Tyr Trp Met Arg His  Ala Leu Thr
        995                 1000                1005

Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu Val Pro  Val Gly Ala
    1010                1015                1020

Val Leu  Val Leu Asn Asn Arg  Val Ile Gly Glu Gly  Trp Asn Arg
    1025                1030                1035

Ala Ile  Gly Leu His Asp Pro  Thr Ala His Ala Glu  Ile Met Ala
    1040                1045                1050

Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn Tyr Arg  Leu Ile Asp
    1055                1060                1065

Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys Val Met  Cys Ala Gly
    1070                1075                1080

Ala Met  Ile His Ser Arg Ile  Gly Arg Val Val Phe  Gly Val Arg
    1085                1090                1095
```

```
Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu Met Asp  Val Leu His
    1100                1105                1110

Tyr Pro  Gly Met Asn His Arg  Val Glu Ile Thr Glu  Gly Ile Leu
    1115                1120                1125

Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro
    1130                1135                1140

Arg Gln  Val Phe Asn Ala Gln  Lys Lys Ala Gln Ser  Ser Thr Asp
    1145                1150                1155

Gly Ser  Ser Gly Ser Ser Gly  Ser Glu Thr Pro Gly  Thr Ser Glu
    1160                1165                1170

Ser Ala  Thr Pro Glu Ser Ser  Gly Gly Ser Ser Ile  Lys Lys Tyr
    1175                1180                1185

Pro Lys  Leu Glu Ser Glu Phe  Val Tyr Gly Asp Tyr  Lys Val Tyr
    1190                1195                1200

Asp Val  Arg Lys Met Ile Ala  Lys Ser Glu Gln Glu  Ile Gly Lys
    1205                1210                1215

Ala Thr  Ala Lys Tyr Phe Phe  Tyr Ser Asn Ile Met  Asn Phe Phe
    1220                1225                1230

Lys Thr  Glu Ile Thr Leu Ala  Asn Gly Glu Ile Arg  Lys Arg Pro
    1235                1240                1245

Leu Ile  Glu Thr Asn Gly Glu  Thr Gly Glu Ile Val  Trp Asp Lys
    1250                1255                1260

Gly Arg  Asp Phe Ala Thr Val  Arg Lys Val Leu Ser  Met Pro Gln
    1265                1270                1275

Val Asn  Ile Val Lys Lys Thr  Glu Val Gln Thr Gly  Gly Phe Ser
    1280                1285                1290
```

```
Lys Glu  Ser Ile Leu Pro Lys  Arg Asn Ser Asp Lys  Leu Ile Ala
    1295                 1300                 1305

Arg Lys  Lys Asp Trp Asp Pro  Lys Lys Tyr Gly Gly  Phe Asp Ser
    1310                 1315                 1320

Pro Thr  Val Ala Tyr Ser Val  Leu Val Val Ala Lys  Val Glu Lys
    1325                 1330                 1335

Gly Lys  Ser Lys Lys Leu Lys  Ser Val Lys Glu Leu  Leu Gly Ile
    1340                 1345                 1350

Thr Ile  Met Glu Arg Ser Ser  Phe Glu Lys Asn Pro  Ile Asp Phe
    1355                 1360                 1365

Leu Glu  Ala Lys Gly Tyr Lys  Glu Val Lys Lys Asp  Leu Ile Ile
    1370                 1375                 1380

Lys Leu  Pro Lys Tyr Ser Leu  Phe Glu Leu Glu Asn  Gly Arg Lys
    1385                 1390                 1395

Arg Met  Leu Ala Ser Ala Gly  Glu Leu Gln Lys Gly  Asn Glu Leu
    1400                 1405                 1410

Ala Leu  Pro Ser Lys Tyr Val  Asn Phe Leu Tyr Leu  Ala Ser His
    1415                 1420                 1425

Tyr Glu  Lys Leu Lys Gly Ser  Pro Glu Asp Asn Glu  Gln Lys Gln
    1430                 1435                 1440

Leu Phe  Val Glu Gln His Lys  His Tyr Leu Asp Glu  Ile Ile Glu
    1445                 1450                 1455

Gln Ile  Ser Glu Phe Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn
    1460                 1465                 1470

Leu Asp  Lys Val Leu Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro
    1475                 1480                 1485

Ile Arg  Glu Gln Ala Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr
    1490                 1495                 1500
```

Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile
    1505   1510   1515

Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr
    1520   1525   1530

Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
    1535   1540   1545

Leu Ser Gln Leu Gly Gly Asp
    1550   1555

<210> 283
<211> 1421
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 112 delta HNH TadA
       polypeptide sequence

<400> 283

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

```
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
```

```
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
```

```
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
```

```
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu
785                 790                 795                 800

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu
                805                 810                 815

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile
            820                 825                 830

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
        835                 840                 845

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
    850                 855                 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
865                 870                 875                 880

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
                885                 890                 895

Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
            900                 905                 910

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
        915                 920                 925

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
    930                 935                 940

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Gly
945                 950                 955                 960

Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu
```

```
               965                 970                 975

Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met
            980                 985                 990

Asn Thr Lys Tyr Asp Glu Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val
        995                 1000                1005

Ile Thr  Leu Lys Ser Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe
    1010                1015                1020

Gln Phe  Tyr Lys Val Arg Glu  Ile Asn Asn Tyr His  His Ala His
    1025                1030                1035

Asp Ala  Tyr Leu Asn Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys
    1040                1045                1050

Tyr Pro  Lys Leu Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1055                1060                1065

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1070                1075                1080

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1085                1090                1095

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1100                1105                1110

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1115                1120                1125

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1130                1135                1140

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1145                1150                1155

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1160                1165                1170
```

```
Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1175                1180                1185

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1190                1195                1200

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1205                1210                1215

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1220                1225                1230

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1235                1240                1245

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1250                1255                1260

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1265                1270                1275

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1280                1285                1290

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1295                1300                1305

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
    1310                1315                1320

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
    1325                1330                1335

Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1340                1345                1350

Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1355                1360                1365

Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
```

1370 1375 1380

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1385 1390 1395

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1400 1405 1410

Asp Leu Ser Gln Leu Gly Gly Asp
1415 1420

<210> 284
<211> 1552
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 113 N-term single TadA helix trunc 165-end polypeptide sequence

<400> 284

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 Ala Lys Thr Gly
100 105 110

```
Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Tyr Phe Phe Arg Met Pro Arg Ser Gly Gly Ser Ser Gly Gly
145                 150                 155                 160

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
                165                 170                 175

Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly
            180                 185                 190

Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu
        195                 200                 205

Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg
    210                 215                 220

His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly
225                 230                 235                 240

Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr
                245                 250                 255

Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn
            260                 265                 270

Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser
        275                 280                 285

Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly
    290                 295                 300

Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr
305                 310                 315                 320

His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg
                325                 330                 335
```

```
Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe
            340                 345                 350

Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu
        355                 360                 365

Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro
    370                 375                 380

Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu
385                 390                 395                 400

Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu
                405                 410                 415

Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu
            420                 425                 430

Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu
        435                 440                 445

Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala
    450                 455                 460

Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu
465                 470                 475                 480

Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile
                485                 490                 495

Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His
            500                 505                 510

His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro
        515                 520                 525

Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala
    530                 535                 540
```

```
Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile
545                 550                 555                 560

Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys
                565                 570                 575

Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly
            580                 585                 590

Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg
        595                 600                 605

Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile
    610                 615                 620

Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala
625                 630                 635                 640

Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr
                645                 650                 655

Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala
            660                 665                 670

Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn
        675                 680                 685

Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val
    690                 695                 700

Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys
705                 710                 715                 720

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

Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr
            740                 745                 750

Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu
        755                 760                 765
```

```
Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile
    770                 775                 780

Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu
785                 790                 795                 800

Glu Asp Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile
                805                 810                 815

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

Lys Gln Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg
        835                 840                 845

Lys Leu Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu
    850                 855                 860

Asp Phe Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu
865                 870                 875                 880

Ile His Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln
                885                 890                 895

Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala
            900                 905                 910

Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val
        915                 920                 925

Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val
    930                 935                 940

Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn
945                 950                 955                 960

Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly
                965                 970                 975
```

```
Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn
            980                 985                 990

Glu Lys Leu Tyr Leu Tyr Tyr Leu  Gln Asn Gly Arg Asp  Met Tyr Val
        995                 1000                1005

Asp Gln  Glu Leu Asp Ile Asn  Arg Leu Ser Asp Tyr  Asp Val Asp
    1010                1015                1020

His Ile  Val Pro Gln Ser Phe  Leu Lys Asp Asp Ser  Ile Asp Asn
    1025                1030                1035

Lys Val  Leu Thr Arg Ser Asp  Lys Asn Arg Gly Lys  Ser Asp Asn
    1040                1045                1050

Val Pro  Ser Glu Glu Val Val  Lys Lys Met Lys Asn  Tyr Trp Arg
    1055                1060                1065

Gln Leu  Leu Asn Ala Lys Leu  Ile Thr Gln Arg Lys  Phe Asp Asn
    1070                1075                1080

Leu Thr  Lys Ala Glu Arg Gly  Gly Leu Ser Glu Leu  Asp Lys Ala
    1085                1090                1095

Gly Phe  Ile Lys Arg Gln Leu  Val Glu Thr Arg Gln  Ile Thr Lys
    1100                1105                1110

His Val  Ala Gln Ile Leu Asp  Ser Arg Met Asn Thr  Lys Tyr Asp
    1115                1120                1125

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

Ser Lys  Leu Val Ser Asp Phe  Arg Lys Asp Phe Gln  Phe Tyr Lys
    1145                1150                1155

Val Arg  Glu Ile Asn Asn Tyr  His His Ala His Asp  Ala Tyr Leu
    1160                1165                1170

Asn Ala  Val Val Gly Thr Ala  Leu Ile Lys Lys Tyr  Pro Lys Leu
    1175                1180                1185
```

```
Glu Ser  Glu Phe Val Tyr Gly  Asp Tyr Lys Val Tyr  Asp Val Arg
    1190                1195                1200

Lys Met  Ile Ala Lys Ser Glu  Gln Glu Ile Gly Lys  Ala Thr Ala
    1205                1210                1215

Lys Tyr  Phe Phe Tyr Ser Asn  Ile Met Asn Phe Phe  Lys Thr Glu
    1220                1225                1230

Ile Thr  Leu Ala Asn Gly Glu  Ile Arg Lys Arg Pro  Leu Ile Glu
    1235                1240                1245

Thr Asn  Gly Glu Thr Gly Glu  Ile Val Trp Asp Lys  Gly Arg Asp
    1250                1255                1260

Phe Ala  Thr Val Arg Lys Val  Leu Ser Met Pro Gln  Val Asn Ile
    1265                1270                1275

Val Lys  Lys Thr Glu Val Gln  Thr Gly Gly Phe Ser  Lys Glu Ser
    1280                1285                1290

Ile Leu  Pro Lys Arg Asn Ser  Asp Lys Leu Ile Ala  Arg Lys Lys
    1295                1300                1305

Asp Trp  Asp Pro Lys Lys Tyr  Gly Gly Phe Asp Ser  Pro Thr Val
    1310                1315                1320

Ala Tyr  Ser Val Leu Val Val  Ala Lys Val Glu Lys  Gly Lys Ser
    1325                1330                1335

Lys Lys  Leu Lys Ser Val Lys  Glu Leu Leu Gly Ile  Thr Ile Met
    1340                1345                1350

Glu Arg  Ser Ser Phe Glu Lys  Asn Pro Ile Asp Phe  Leu Glu Ala
    1355                1360                1365

Lys Gly  Tyr Lys Glu Val Lys  Lys Asp Leu Ile Ile  Lys Leu Pro
    1370                1375                1380
```

```
Lys Tyr  Ser Leu Phe Glu Leu  Glu Asn Gly Arg Lys  Arg Met Leu
    1385                1390                1395


Ala Ser  Ala Gly Glu Leu Gln  Lys Gly Asn Glu Leu  Ala Leu Pro
    1400                1405                1410


Ser Lys  Tyr Val Asn Phe Leu  Tyr Leu Ala Ser His  Tyr Glu Lys
    1415                1420                1425


Leu Lys  Gly Ser Pro Glu Asp  Asn Glu Gln Lys Gln  Leu Phe Val
    1430                1435                1440


Glu Gln  His Lys His Tyr Leu  Asp Glu Ile Ile Glu  Gln Ile Ser
    1445                1450                1455


Glu Phe  Ser Lys Arg Val Ile  Leu Ala Asp Ala Asn  Leu Asp Lys
    1460                1465                1470


Val Leu  Ser Ala Tyr Asn Lys  His Arg Asp Lys Pro  Ile Arg Glu
    1475                1480                1485


Gln Ala  Glu Asn Ile Ile His  Leu Phe Thr Leu Thr  Asn Leu Gly
    1490                1495                1500


Ala Pro  Ala Ala Phe Lys Tyr  Phe Asp Thr Thr Ile  Asp Arg Lys
    1505                1510                1515


Arg Tyr  Thr Ser Thr Lys Glu  Val Leu Asp Ala Thr  Leu Ile His
    1520                1525                1530


Gln Ser  Ile Thr Gly Leu Tyr  Glu Thr Arg Ile Asp  Leu Ser Gln
    1535                1540                1545


Leu Gly  Gly Asp
    1550
```

<210> 285
<211> 1545
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of artificial sequence: 114 N-term single TadA helix trunc 165-end delta 59-65 polypeptide sequence

<400> 285

```
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 Thr
        35                  40                  45

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln
    50                  55                  60

Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
65                  70                  75                  80

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
                85                  90                  95

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
            100                 105                 110

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly
        115                 120                 125

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met
    130                 135                 140

Pro Arg Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro
145                 150                 155                 160

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly
                165                 170                 175

Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser
            180                 185                 190

Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys
```

```
        195                 200                 205

Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu
    210                 215                 220

Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg
225                 230                 235                 240

Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile
                245                 250                 255

Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp
            260                 265                 270

Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys
        275                 280                 285

Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala
    290                 295                 300

Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val
305                 310                 315                 320

Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala
                325                 330                 335

His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn
            340                 345                 350

Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr
        355                 360                 365

Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp
    370                 375                 380

Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu
385                 390                 395                 400

Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly
                405                 410                 415
```

```
Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn
            420                 425                 430

Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr
        435                 440                 445

Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala
    450                 455                 460

Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser
465                 470                 475                 480

Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala
                485                 490                 495

Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu
            500                 505                 510

Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe
        515                 520                 525

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

Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met
545                 550                 555                 560

Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu
                565                 570                 575

Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His
            580                 585                 590

Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro
        595                 600                 605

Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg
    610                 615                 620

Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala
```

625 630 635 640

Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu
645 650 655

Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met
660 665 670

Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His
675 680 685

Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val
690 695 700

Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu
705 710 715 720

Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val
725 730 735

Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe
740 745 750

Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu
755 760 765

Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu
770 775 780

Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu
785 790 795 800

Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr
805 810 815

Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg
820 825 830

Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg
835 840 845

```
Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
    850                 855                 860

Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr
865                 870                 875                 880

Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser
                885                 890                 895

Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys
            900                 905                 910

Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met
        915                 920                 925

Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn
    930                 935                 940

Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg
945                 950                 955                 960

Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His
                965                 970                 975

Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr
            980                 985                 990

Leu Gln Asn Gly Arg Asp Met Tyr  Val Asp Gln Glu Leu  Asp Ile Asn
        995                 1000                1005

Arg Leu  Ser Asp Tyr Asp Val  Asp His Ile Val Pro  Gln Ser Phe
    1010                1015                1020

Leu Lys  Asp Asp Ser Ile Asp  Asn Lys Val Leu Thr  Arg Ser Asp
    1025                1030                1035

Lys Asn  Arg Gly Lys Ser Asp  Asn Val Pro Ser Glu  Glu Val Val
    1040                1045                1050

Lys Lys  Met Lys Asn Tyr Trp  Arg Gln Leu Leu Asn  Ala Lys Leu
```

```
    1055                1060                1065

Ile Thr  Gln Arg Lys Phe Asp  Asn Leu Thr Lys Ala  Glu Arg Gly
    1070                1075                1080

Gly Leu  Ser Glu Leu Asp Lys  Ala Gly Phe Ile Lys  Arg Gln Leu
    1085                1090                1095

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

Ser Arg  Met Asn Thr Lys Tyr  Asp Glu Asn Asp Lys  Leu Ile Arg
    1115                1120                1125

Glu Val  Lys Val Ile Thr Leu  Lys Ser Lys Leu Val  Ser Asp Phe
    1130                1135                1140

Arg Lys  Asp Phe Gln Phe Tyr  Lys Val Arg Glu Ile  Asn Asn Tyr
    1145                1150                1155

His His  Ala His Asp Ala Tyr  Leu Asn Ala Val Val  Gly Thr Ala
    1160                1165                1170

Leu Ile  Lys Lys Tyr Pro Lys  Leu Glu Ser Glu Phe  Val Tyr Gly
    1175                1180                1185

Asp Tyr  Lys Val Tyr Asp Val  Arg Lys Met Ile Ala  Lys Ser Glu
    1190                1195                1200

Gln Glu  Ile Gly Lys Ala Thr  Ala Lys Tyr Phe Phe  Tyr Ser Asn
    1205                1210                1215

Ile Met  Asn Phe Phe Lys Thr  Glu Ile Thr Leu Ala  Asn Gly Glu
    1220                1225                1230

Ile Arg  Lys Arg Pro Leu Ile  Glu Thr Asn Gly Glu  Thr Gly Glu
    1235                1240                1245

Ile Val  Trp Asp Lys Gly Arg  Asp Phe Ala Thr Val  Arg Lys Val
    1250                1255                1260
```

```
Leu Ser  Met Pro Gln Val Asn  Ile Val Lys Lys Thr  Glu Val Gln
    1265                1270                1275

Thr Gly  Gly Phe Ser Lys Glu  Ser Ile Leu Pro Lys  Arg Asn Ser
    1280                1285                1290

Asp Lys  Leu Ile Ala Arg Lys  Lys Asp Trp Asp Pro  Lys Lys Tyr
    1295                1300                1305

Gly Gly  Phe Asp Ser Pro Thr  Val Ala Tyr Ser Val  Leu Val Val
    1310                1315                1320

Ala Lys  Val Glu Lys Gly Lys  Ser Lys Lys Leu Lys  Ser Val Lys
    1325                1330                1335

Glu Leu  Leu Gly Ile Thr Ile  Met Glu Arg Ser Ser  Phe Glu Lys
    1340                1345                1350

Asn Pro  Ile Asp Phe Leu Glu  Ala Lys Gly Tyr Lys  Glu Val Lys
    1355                1360                1365

Lys Asp  Leu Ile Ile Lys Leu  Pro Lys Tyr Ser Leu  Phe Glu Leu
    1370                1375                1380

Glu Asn  Gly Arg Lys Arg Met  Leu Ala Ser Ala Gly  Glu Leu Gln
    1385                1390                1395

Lys Gly  Asn Glu Leu Ala Leu  Pro Ser Lys Tyr Val  Asn Phe Leu
    1400                1405                1410

Tyr Leu  Ala Ser His Tyr Glu  Lys Leu Lys Gly Ser  Pro Glu Asp
    1415                1420                1425

Asn Glu  Gln Lys Gln Leu Phe  Val Glu Gln His Lys  His Tyr Leu
    1430                1435                1440

Asp Glu  Ile Ile Glu Gln Ile  Ser Glu Phe Ser Lys  Arg Val Ile
    1445                1450                1455

Leu Ala  Asp Ala Asn Leu Asp  Lys Val Leu Ser Ala  Tyr Asn Lys
```

```
    1460                1465                1470

His Arg  Asp Lys Pro Ile Arg  Glu Gln Ala Glu Asn  Ile Ile His
    1475                1480                1485

Leu Phe  Thr Leu Thr Asn Leu  Gly Ala Pro Ala Ala  Phe Lys Tyr
    1490                1495                1500

Phe Asp  Thr Thr Ile Asp Arg  Lys Arg Tyr Thr Ser  Thr Lys Glu
    1505                1510                1515

Val Leu  Asp Ala Thr Leu Ile  His Gln Ser Ile Thr  Gly Leu Tyr
    1520                1525                1530

Glu Thr  Arg Ile Asp Leu Ser  Gln Leu Gly Gly Asp
    1535                1540                1545
```

<210> 286
<211> 1541
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 115.1 Cas9 TadAins1004 polypeptide sequence

<400> 286

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80
```

```
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300
```

```
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
```

```
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735
```

```
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
```

```
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Gly Ser  Ser Gly Ser
        995                 1000                1005


Glu Thr  Pro Gly Thr Ser Glu  Ser Ala Thr Pro Glu  Ser Ser Gly
    1010                1015                1020


Ser Glu  Val Glu Phe Ser His  Glu Tyr Trp Met Arg  His Ala Leu
    1025                1030                1035


Thr Leu  Ala Lys Arg Ala Arg  Asp Glu Arg Glu Val  Pro Val Gly
    1040                1045                1050


Ala Val  Leu Val Leu Asn Asn  Arg Val Ile Gly Glu  Gly Trp Asn
    1055                1060                1065


Arg Ala  Ile Gly Leu His Asp  Pro Thr Ala His Ala  Glu Ile Met
    1070                1075                1080


Ala Leu  Arg Gln Gly Gly Leu  Val Met Gln Asn Tyr  Arg Leu Ile
    1085                1090                1095


Asp Ala  Thr Leu Tyr Val Thr  Phe Glu Pro Cys Val  Met Cys Ala
    1100                1105                1110


Gly Ala  Met Ile His Ser Arg  Ile Gly Arg Val Val  Phe Gly Val
    1115                1120                1125


Arg Asn  Ala Lys Thr Gly Ala  Ala Gly Ser Leu Met  Asp Val Leu
    1130                1135                1140


His Tyr  Pro Gly Met Asn His  Arg Val Glu Ile Thr  Glu Gly Ile
    1145                1150                1155
```

```
Leu Ala  Asp Glu Cys Ala Ala  Leu Leu Cys Tyr Phe  Phe Arg Met
    1160                1165                1170

Pro Arg  Gln Leu Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                1255                1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                1270                1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                1285                1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                1300                1305

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                1315                1320

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                1330                1335

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                1345                1350
```

```
Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                1360                1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                1375                1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                1390                1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                1405                1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1415                1420                1425

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
    1430                1435                1440

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
    1445                1450                1455

Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1460                1465                1470

Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1475                1480                1485

Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
    1490                1495                1500

Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala
    1505                1510                1515

Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile
    1520                1525                1530

Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540
```

<210> 287
<211> 1541

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 115.2 Cas9 TadAins1005 polypeptide sequence

<400> 287

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

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

```
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400
```

```
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
```

```
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
```

```
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Gly  Ser Ser Gly
        995                 1000                1005

Ser Glu  Thr Pro Gly Thr Ser  Glu Ser Ala Thr Pro  Glu Ser Ser
    1010                1015                1020

Gly Ser  Glu Val Glu Phe Ser  His Glu Tyr Trp Met  Arg His Ala
    1025                1030                1035

Leu Thr  Leu Ala Lys Arg Ala  Arg Asp Glu Arg Glu  Val Pro Val
```

```
    1040                1045                1050

Gly Ala  Val Leu Val Leu Asn  Asn Arg Val Ile Gly  Glu Gly Trp
    1055                1060                1065

Asn Arg  Ala Ile Gly Leu His  Asp Pro Thr Ala His  Ala Glu Ile
    1070                1075                1080

Met Ala  Leu Arg Gln Gly Gly  Leu Val Met Gln Asn  Tyr Arg Leu
    1085                1090                1095

Ile Asp  Ala Thr Leu Tyr Val  Thr Phe Glu Pro Cys  Val Met Cys
    1100                1105                1110

Ala Gly  Ala Met Ile His Ser  Arg Ile Gly Arg Val  Val Phe Gly
    1115                1120                1125

Val Arg  Asn Ala Lys Thr Gly  Ala Ala Gly Ser Leu  Met Asp Val
    1130                1135                1140

Leu His  Tyr Pro Gly Met Asn  His Arg Val Glu Ile  Thr Glu Gly
    1145                1150                1155

Ile Leu  Ala Asp Glu Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg
    1160                1165                1170

Met Pro  Arg Gln Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245
```

```
Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                 1255                 1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                 1270                 1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                 1285                 1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                 1300                 1305

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                 1315                 1320

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                 1330                 1335

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                 1345                 1350

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                 1360                 1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                 1375                 1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                 1390                 1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                 1405                 1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1415                 1420                 1425

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
    1430                 1435                 1440

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
```

```
    1445                1450                1455


Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1460                1465                1470


Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1475                1480                1485


Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
    1490                1495                1500


Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala
    1505                1510                1515


Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile
    1520                1525                1530


Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540


<210>  288
<211>  1541
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 115.3 Cas9 TadAins1006
       polypeptide sequence

<400>  288

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60
```

```
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
```

```
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
```

```
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720
```

```
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
               725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
               805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
               885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925
```

```
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Gly Ser Ser
        995                 1000                1005

Gly Ser  Glu Thr Pro Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser
    1010                1015                1020

Ser Gly  Ser Glu Val Glu Phe  Ser His Glu Tyr Trp  Met Arg His
    1025                1030                1035

Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro
    1040                1045                1050

Val Gly  Ala Val Leu Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly
    1055                1060                1065

Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro Thr Ala  His Ala Glu
    1070                1075                1080

Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg
    1085                1090                1095

Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met
    1100                1105                1110

Cys Ala  Gly Ala Met Ile His  Ser Arg Ile Gly Arg  Val Val Phe
    1115                1120                1125

Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp
    1130                1135                1140
```

```
Val Leu  His Tyr Pro Gly Met  Asn His Arg Val Glu  Ile Thr Glu
    1145                1150                1155

Gly Ile  Leu Ala Asp Glu Cys  Ala Ala Leu Leu Cys  Tyr Phe Phe
    1160                1165                1170

Arg Met  Pro Arg Gln Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                1255                1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                1270                1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                1285                1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                1300                1305

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                1315                1320

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                1330                1335
```

```
Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                1345                1350

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                1360                1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                1375                1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                1390                1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                1405                1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1415                1420                1425

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
    1430                1435                1440

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
    1445                1450                1455

Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1460                1465                1470

Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1475                1480                1485

Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
    1490                1495                1500

Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala
    1505                1510                1515

Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile
    1520                1525                1530

Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540
```

<210> 289
<211> 1541
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 115.4 Cas9 TadAins1007 polypeptide sequence

<400> 289

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

```
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380
```

```
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
```

```
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815
```

```
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Gly Ser
        995                 1000                1005

Ser Gly  Ser Glu Thr Pro Gly  Thr Ser Glu Ser Ala  Thr Pro Glu
    1010                1015                1020

Ser Ser  Gly Ser Glu Val Glu  Phe Ser His Glu Tyr  Trp Met Arg
```

```
    1025                1030                1035

His Ala  Leu Thr Leu Ala Lys  Arg Ala Arg Asp Glu  Arg Glu Val
    1040                1045                1050

Pro Val  Gly Ala Val Leu Val  Leu Asn Asn Arg Val  Ile Gly Glu
    1055                1060                1065

Gly Trp  Asn Arg Ala Ile Gly  Leu His Asp Pro Thr  Ala His Ala
    1070                1075                1080

Glu Ile  Met Ala Leu Arg Gln  Gly Gly Leu Val Met  Gln Asn Tyr
    1085                1090                1095

Arg Leu  Ile Asp Ala Thr Leu  Tyr Val Thr Phe Glu  Pro Cys Val
    1100                1105                1110

Met Cys  Ala Gly Ala Met Ile  His Ser Arg Ile Gly  Arg Val Val
    1115                1120                1125

Phe Gly  Val Arg Asn Ala Lys  Thr Gly Ala Ala Gly  Ser Leu Met
    1130                1135                1140

Asp Val  Leu His Tyr Pro Gly  Met Asn His Arg Val  Glu Ile Thr
    1145                1150                1155

Glu Gly  Ile Leu Ala Asp Glu  Cys Ala Ala Leu Leu  Cys Tyr Phe
    1160                1165                1170

Phe Arg  Met Pro Arg Gln Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230
```

```
Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                1255                1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                1270                1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                1285                1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                1300                1305

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                1315                1320

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                1330                1335

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                1345                1350

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                1360                1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                1375                1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                1390                1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                1405                1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1415                1420                1425

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
```

```
    1430                1435                1440

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
    1445                1450                1455

Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1460                1465                1470

Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1475                1480                1485

Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
    1490                1495                1500

Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala
    1505                1510                1515

Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile
    1520                1525                1530

Asp Leu  Ser Gln Leu Gly Gly  Asp
    1535                1540
```

<210> 290
<211> 1541
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 116.1 Cas9 TadAins C-term truncate2 792 polypeptide sequence

<400> 290

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
```

```
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
```

```
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480
```

```
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

```
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Gly Ser Ser Gly Ser Glu Thr Pro
785                 790                 795                 800

Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu
                805                 810                 815

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
            820                 825                 830

Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn
        835                 840                 845

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp
    850                 855                 860

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
865                 870                 875                 880

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu
                885                 890                 895

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
            900                 905                 910
```

```
Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu
        915                 920                 925


Met Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr
    930                 935                 940


Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe
945                 950                 955                 960


Arg Met Pro Arg Gln Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
                965                 970                 975


Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
            980                 985                 990


Arg Asp Met Tyr Val Asp Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp
        995                 1000                1005


Tyr Asp  Val Asp His Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp
    1010                1015                1020


Ser Ile  Asp Asn Lys Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly
    1025                1030                1035


Lys Ser  Asp Asn Val Pro Ser  Glu Glu Val Val Lys  Lys Met Lys
    1040                1045                1050


Asn Tyr  Trp Arg Gln Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg
    1055                1060                1065


Lys Phe  Asp Asn Leu Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu
    1070                1075                1080


Leu Asp  Lys Ala Gly Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg
    1085                1090                1095


Gln Ile  Thr Lys His Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn
    1100                1105                1110


Thr Lys  Tyr Asp Glu Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val
    1115                1120                1125
```

```
Ile Thr  Leu Lys Ser Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe
    1130                1135                1140

Gln Phe  Tyr Lys Val Arg Glu  Ile Asn Asn Tyr His  His Ala His
    1145                1150                1155

Asp Ala  Tyr Leu Asn Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys
    1160                1165                1170

Tyr Pro  Lys Leu Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                1255                1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                1270                1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                1285                1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                1300                1305

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                1315                1320
```

```
Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                1330                1335

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                1345                1350

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                1360                1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                1375                1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                1390                1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                1405                1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1415                1420                1425

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
    1430                1435                1440

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
    1445                1450                1455

Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1460                1465                1470

Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1475                1480                1485

Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
    1490                1495                1500

Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala
    1505                1510                1515

Thr Leu  Ile His Gln Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile
    1520                1525                1530
```

Asp Leu Ser Gln Leu Gly Gly Asp
    1535   1540

<210> 291
<211> 1541
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 116.2 Cas9 TadAins C-term
       truncate2 791 polypeptide sequence

<400> 291

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

```
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
```

```
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Ser Gly Ser Glu Thr Pro Gly
785                 790                 795                 800
```

```
Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe
                805                 810                 815

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala
            820                 825                 830

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn
        835                 840                 845

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
    850                 855                 860

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
865                 870                 875                 880

Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
                885                 890                 895

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
            900                 905                 910

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
        915                 920                 925

Asp Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
    930                 935                 940

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg
945                 950                 955                 960

Met Pro Arg Gln Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
                965                 970                 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
            980                 985                 990

Arg Asp Met Tyr Val Asp Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp
        995                 1000                1005

Tyr Asp  Val Asp His Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp
```

```
    1010                1015                1020

Ser Ile  Asp Asn Lys Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly
    1025                1030                1035

Lys Ser  Asp Asn Val Pro Ser  Glu Glu Val Val Lys  Lys Met Lys
    1040                1045                1050

Asn Tyr  Trp Arg Gln Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg
    1055                1060                1065

Lys Phe  Asp Asn Leu Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu
    1070                1075                1080

Leu Asp  Lys Ala Gly Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg
    1085                1090                1095

Gln Ile  Thr Lys His Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn
    1100                1105                1110

Thr Lys  Tyr Asp Glu Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val
    1115                1120                1125

Ile Thr  Leu Lys Ser Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe
    1130                1135                1140

Gln Phe  Tyr Lys Val Arg Glu  Ile Asn Asn Tyr His  His Ala His
    1145                1150                1155

Asp Ala  Tyr Leu Asn Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys
    1160                1165                1170

Tyr Pro  Lys Leu Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215
```

```
Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                1255                1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                1270                1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                1285                1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                1300                1305

Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                1315                1320

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                1330                1335

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                1345                1350

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                1360                1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                1375                1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                1390                1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                1405                1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
```

1415 1420 1425

Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
1430 1435 1440

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala
1445 1450 1455

Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys
1460 1465 1470

Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu
1475 1480 1485

Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr
1490 1495 1500

Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1505 1510 1515

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1520 1525 1530

Asp Leu Ser Gln Leu Gly Gly Asp
1535 1540

<210> 292
<211> 1541
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 116.3 Cas9 TadAins C-term truncate2 790 polypeptide sequence

<400> 292

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

```
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255
```

```
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460
```

```
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
```

```
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
785                 790                 795                 800

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser
                805                 810                 815

His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg
            820                 825                 830

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg
        835                 840                 845

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr
    850                 855                 860

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln
865                 870                 875                 880

Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
                885                 890                 895
```

```
Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
            900                 905                 910

Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
        915                 920                 925

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly
    930                 935                 940

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met
945                 950                 955                 960

Pro Arg Gln Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn
                965                 970                 975

Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly
            980                 985                 990

Arg Asp Met Tyr Val Asp Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp
        995                 1000                1005

Tyr Asp  Val Asp His Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp
    1010                1015                1020

Ser Ile  Asp Asn Lys Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly
    1025                1030                1035

Lys Ser  Asp Asn Val Pro Ser  Glu Glu Val Val Lys  Lys Met Lys
    1040                1045                1050

Asn Tyr  Trp Arg Gln Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg
    1055                1060                1065

Lys Phe  Asp Asn Leu Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu
    1070                1075                1080

Leu Asp  Lys Ala Gly Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg
    1085                1090                1095

Gln Ile  Thr Lys His Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn
    1100                1105                1110
```

```
Thr Lys  Tyr Asp Glu Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val
    1115                1120                1125

Ile Thr  Leu Lys Ser Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe
    1130                1135                1140

Gln Phe  Tyr Lys Val Arg Glu  Ile Asn Asn Tyr His  His Ala His
    1145                1150                1155

Asp Ala  Tyr Leu Asn Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys
    1160                1165                1170

Tyr Pro  Lys Leu Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1175                1180                1185

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly
    1190                1195                1200

Lys Ala  Thr Ala Lys Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe
    1205                1210                1215

Phe Lys  Thr Glu Ile Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg
    1220                1225                1230

Pro Leu  Ile Glu Thr Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp
    1235                1240                1245

Lys Gly  Arg Asp Phe Ala Thr  Val Arg Lys Val Leu  Ser Met Pro
    1250                1255                1260

Gln Val  Asn Ile Val Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe
    1265                1270                1275

Ser Lys  Glu Ser Ile Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile
    1280                1285                1290

Ala Arg  Lys Lys Asp Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp
    1295                1300                1305
```

```
Ser Pro  Thr Val Ala Tyr Ser  Val Leu Val Val Ala  Lys Val Glu
    1310                1315                1320

Lys Gly  Lys Ser Lys Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly
    1325                1330                1335

Ile Thr  Ile Met Glu Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp
    1340                1345                1350

Phe Leu  Glu Ala Lys Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile
    1355                1360                1365

Ile Lys  Leu Pro Lys Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg
    1370                1375                1380

Lys Arg  Met Leu Ala Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu
    1385                1390                1395

Leu Ala  Leu Pro Ser Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser
    1400                1405                1410

His Tyr  Glu Lys Leu Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys
    1415                1420                1425

Gln Leu  Phe Val Glu Gln His  Lys His Tyr Leu Asp  Glu Ile Ile
    1430                1435                1440

Glu Gln  Ile Ser Glu Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala
    1445                1450                1455

Asn Leu  Asp Lys Val Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys
    1460                1465                1470

Pro Ile  Arg Glu Gln Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu
    1475                1480                1485

Thr Asn  Leu Gly Ala Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr
    1490                1495                1500

Ile Asp  Arg Lys Arg Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala
    1505                1510                1515
```

Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
    1520   1525   1530

Asp Leu Ser Gln Leu Gly Gly Asp
    1535   1540

<210> 293
<211> 1484
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 117 Cas9 delta 1017-1069
       polypeptide sequence

<400> 293

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

```
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
```

```
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
```

```
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
```

```
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800


Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815


Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830


Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845


Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860


Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880


Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895


Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910


Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925


Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940


Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
```

```
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Ser Ser Gly Ser  Glu Val Glu
    1010                1015                1020

Phe Ser  His Glu Tyr Trp Met  Arg His Ala Leu Thr  Leu Ala Lys
    1025                1030                1035

Arg Ala  Arg Asp Glu Arg Glu  Val Pro Val Gly Ala  Val Leu Val
    1040                1045                1050

Leu Asn  Asn Arg Val Ile Gly  Glu Gly Trp Asn Arg  Ala Ile Gly
    1055                1060                1065

Leu His  Asp Pro Thr Ala His  Ala Glu Ile Met Ala  Leu Arg Gln
    1070                1075                1080

Gly Gly  Leu Val Met Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu
    1085                1090                1095

Tyr Val  Thr Phe Glu Pro Cys  Val Met Cys Ala Gly  Ala Met Ile
    1100                1105                1110

His Ser  Arg Ile Gly Arg Val  Val Phe Gly Val Arg  Asn Ala Lys
    1115                1120                1125

Thr Gly  Ala Ala Gly Ser Leu  Met Asp Val Leu His  Tyr Pro Gly
    1130                1135                1140

Met Asn  His Arg Val Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu
    1145                1150                1155

Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val
    1160                1165                1170

Phe Asn  Ala Gln Lys Lys Ala  Gln Ser Ser Thr Asp  Gly Glu Ile
    1175                1180                1185

Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu
    1190                1195                1200
```

```
Ser Met  Pro Gln Val Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr
    1205                1210                1215

Gly Gly  Phe Ser Lys Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp
    1220                1225                1230

Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly
    1235                1240                1245

Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser Val Leu  Val Val Ala
    1250                1255                1260

Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu
    1265                1270                1275

Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn
    1280                1285                1290

Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys
    1295                1300                1305

Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu
    1310                1315                1320

Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys
    1325                1330                1335

Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr
    1340                1345                1350

Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn
    1355                1360                1365

Glu Gln  Lys Gln Leu Phe Val  Glu Gln His Lys His  Tyr Leu Asp
    1370                1375                1380

Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu
    1385                1390                1395

Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His
```

1400 1405 1410

Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
1415 1420 1425

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe
1430 1435 1440

Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val
1445 1450 1455

Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu
1460 1465 1470

Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1475 1480

<210> 294
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 118 Cas9 TadA-CP116ins 1067 polypeptide sequence

<400> 294

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

```
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300
```

```
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
```

```
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735
```

```
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
```

```
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Met Asn
    1055                1060                1065


His Arg  Val Glu Ile Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala
    1070                1075                1080


Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro Arg Gln  Val Phe Asn
    1085                1090                1095


Ala Gln  Lys Lys Ala Gln Ser  Ser Thr Asp Gly Ser  Ser Gly Ser
    1100                1105                1110


Glu Thr  Pro Gly Thr Ser Glu  Ser Ala Thr Pro Glu  Ser Ser Gly
    1115                1120                1125


Ser Glu  Val Glu Phe Ser His  Glu Tyr Trp Met Arg  His Ala Leu
    1130                1135                1140


Thr Leu  Ala Lys Arg Ala Arg  Asp Glu Arg Glu Val  Pro Val Gly
    1145                1150                1155
```

```
Ala Val  Leu Val Leu Asn Asn  Arg Val Ile Gly Glu  Gly Trp Asn
    1160                1165                1170

Arg Ala  Ile Gly Leu His Asp  Pro Thr Ala His Ala  Glu Ile Met
    1175                1180                1185

Ala Leu  Arg Gln Gly Gly Leu  Val Met Gln Asn Tyr  Arg Leu Ile
    1190                1195                1200

Asp Ala  Thr Leu Tyr Val Thr  Phe Glu Pro Cys Val  Met Cys Ala
    1205                1210                1215

Gly Ala  Met Ile His Ser Arg  Ile Gly Arg Val Val  Phe Gly Val
    1220                1225                1230

Arg Asn  Ala Lys Thr Gly Ala  Ala Gly Ser Leu Met  Asp Val Leu
    1235                1240                1245

His Tyr  Pro Gly Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350
```

```
Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545

Ser Gln  Leu Gly Gly Asp
    1550
```

<210> 295
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 119 Cas9 TadAins 701 polypeptide sequence

<400> 295

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro

```
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380
```

```
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
```

```
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Gly Ser Ser
    690                 695                 700

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
705                 710                 715                 720

Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
                725                 730                 735

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
            740                 745                 750

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
        755                 760                 765

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
    770                 775                 780

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
785                 790                 795                 800

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                805                 810                 815
```

```
Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
            820                 825                 830

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
        835                 840                 845

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    850                 855                 860

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
865                 870                 875                 880

Lys Ala Gln Ser Ser Thr Asp Leu Thr Phe Lys Glu Asp Ile Gln Lys
                885                 890                 895

Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn
            900                 905                 910

Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys
        915                 920                 925

Val Val Asp Glu Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn
    930                 935                 940

Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln
945                 950                 955                 960

Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu
                965                 970                 975

Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu
            980                 985                 990

Gln Asn Glu Lys Leu Tyr Leu Tyr  Tyr Leu Gln Asn Gly  Arg Asp Met
        995                 1000                1005

Tyr Val  Asp Gln Glu Leu Asp  Ile Asn Arg Leu Ser  Asp Tyr Asp
    1010                1015                1020

Val Asp  His Ile Val Pro Gln  Ser Phe Leu Lys Asp  Asp Ser Ile
```

```
    1025                1030                1035

Asp Asn  Lys Val Leu Thr Arg  Ser Asp Lys Asn Arg  Gly Lys Ser
    1040                1045                1050

Asp Asn  Val Pro Ser Glu Glu  Val Val Lys Lys Met  Lys Asn Tyr
    1055                1060                1065

Trp Arg  Gln Leu Leu Asn Ala  Lys Leu Ile Thr Gln  Arg Lys Phe
    1070                1075                1080

Asp Asn  Leu Thr Lys Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp
    1085                1090                1095

Lys Ala  Gly Phe Ile Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile
    1100                1105                1110

Thr Lys  His Val Ala Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys
    1115                1120                1125

Tyr Asp  Glu Asn Asp Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr
    1130                1135                1140

Leu Lys  Ser Lys Leu Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe
    1145                1150                1155

Tyr Lys  Val Arg Glu Ile Asn  Asn Tyr His His Ala  His Asp Ala
    1160                1165                1170

Tyr Leu  Asn Ala Val Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro
    1175                1180                1185

Lys Leu  Glu Ser Glu Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp
    1190                1195                1200

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230
```

```
Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
```

```
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545

Ser Gln  Leu Gly Gly Asp
    1550
```

<210> 296
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 120 Cas9 TadACP136ins 1248 polypeptide sequence

<400> 296

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30
```

```
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255
```

```
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460
```

```
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
```

```
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
```

```
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110
```

```
Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Met Asn  His Arg Val Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu
    1250                1255                1260

Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val
    1265                1270                1275

Phe Asn  Ala Gln Lys Lys Ala  Gln Ser Ser Thr Asp  Gly Ser Ser
    1280                1285                1290

Gly Ser  Glu Thr Pro Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser
    1295                1300                1305
```

```
Ser Gly  Ser Glu Val Glu Phe  Ser His Glu Tyr Trp  Met Arg His
    1310                 1315                 1320

Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro
    1325                 1330                 1335

Val Gly  Ala Val Leu Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly
    1340                 1345                 1350

Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro Thr Ala  His Ala Glu
    1355                 1360                 1365

Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg
    1370                 1375                 1380

Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met
    1385                 1390                 1395

Cys Ala  Gly Ala Met Ile His  Ser Arg Ile Gly Arg  Val Val Phe
    1400                 1405                 1410

Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp
    1415                 1420                 1425

Val Leu  His Tyr Pro Gly Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                 1435                 1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                 1450                 1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                 1465                 1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                 1480                 1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                 1495                 1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                 1510                 1515
```

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
    1520   1525   1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 297
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 121 Cas9 TadACP136ins 1052
       polypeptide sequence

<400> 297

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

```
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335
```

```
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
```

```
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765
```

```
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
```

```
           980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Met Asn  His Arg Val Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu
    1055                1060                1065

Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val
    1070                1075                1080

Phe Asn  Ala Gln Lys Lys Ala  Gln Ser Ser Thr Asp  Gly Ser Ser
    1085                1090                1095

Gly Ser  Glu Thr Pro Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser
    1100                1105                1110

Ser Gly  Ser Glu Val Glu Phe  Ser His Glu Tyr Trp  Met Arg His
    1115                1120                1125

Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro
    1130                1135                1140

Val Gly  Ala Val Leu Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly
    1145                1150                1155

Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro Thr Ala  His Ala Glu
    1160                1165                1170

Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg
    1175                1180                1185
```

```
Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met
    1190                1195                1200

Cys Ala  Gly Ala Met Ile His  Ser Arg Ile Gly Arg  Val Val Phe
    1205                1210                1215

Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp
    1220                1225                1230

Val Leu  His Tyr Pro Gly Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
```

1385 1390 1395

Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 298
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 122 Cas9 TadACP136ins 1041

polypeptide sequence

<400> 298

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45


Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60


Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80


Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95


Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110


His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125


His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140


Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160


Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175


Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190


Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
```

```
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
```

```
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
```

```
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
               645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
               725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
               805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
```

```
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Met Asn His Arg Val  Glu Ile Thr Glu Gly  Ile Leu Ala
    1040                1045                1050

Asp Glu  Cys Ala Ala Leu Leu  Cys Tyr Phe Phe Arg  Met Pro Arg
    1055                1060                1065
```

```
Gln Val  Phe Asn Ala Gln Lys  Lys Ala Gln Ser Ser  Thr Asp Gly
    1070                1075                1080

Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser Glu Ser  Ala Thr Pro
    1085                1090                1095

Glu Ser  Ser Gly Ser Glu Val  Glu Phe Ser His Glu  Tyr Trp Met
    1100                1105                1110

Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu
    1115                1120                1125

Val Pro  Val Gly Ala Val Leu  Val Leu Asn Asn Arg  Val Ile Gly
    1130                1135                1140

Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His Asp Pro  Thr Ala His
    1145                1150                1155

Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn
    1160                1165                1170

Tyr Arg  Leu Ile Asp Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys
    1175                1180                1185

Val Met  Cys Ala Gly Ala Met  Ile His Ser Arg Ile  Gly Arg Val
    1190                1195                1200

Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu
    1205                1210                1215

Met Asp  Val Leu His Tyr Pro  Gly Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260
```

```
Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470
```

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
    1475   1480   1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
    1490   1495   1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
    1505   1510   1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
    1520   1525   1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 299
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 123 Cas9 TadACP139ins 1299
       polypeptide sequence

<400> 299

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

```
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
```

```
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
```

```
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720
```

```
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
```

```
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140
```

```
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                1285                1290

Tyr Asn  Lys His Arg Met Asn  His Arg Val Glu Ile  Thr Glu Gly
    1295                1300                1305

Ile Leu  Ala Asp Glu Cys Ala  Ala Leu Leu Cys Tyr  Phe Phe Arg
    1310                1315                1320

Met Pro  Arg Gln Val Phe Asn  Ala Gln Lys Lys Ala  Gln Ser Ser
    1325                1330                1335

Thr Asp  Gly Ser Ser Gly Ser  Glu Thr Pro Gly Thr  Ser Glu Ser
```

```
    1340                1345                1350

Ala Thr  Pro Glu Ser Ser Gly  Ser Glu Val Glu Phe  Ser His Glu
    1355                1360                1365

Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala Lys Arg  Ala Arg Asp
    1370                1375                1380

Glu Arg  Glu Val Pro Val Gly  Ala Val Leu Val Leu  Asn Asn Arg
    1385                1390                1395

Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile Gly Leu  His Asp Pro
    1400                1405                1410

Thr Ala  His Ala Glu Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val
    1415                1420                1425

Met Gln  Asn Tyr Arg Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe
    1430                1435                1440

Glu Pro  Cys Val Met Cys Ala  Gly Ala Met Ile His  Ser Arg Ile
    1445                1450                1455

Gly Arg  Val Val Phe Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala
    1460                1465                1470

Gly Ser  Leu Met Asp Val Leu  His Tyr Pro Gly Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545
```

Ser Gln Leu Gly Gly Asp
    1550

<210> 300
<211> 1454
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 124 Cas9 delta 792-872
       TadAins polypeptide sequence

<400> 300

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145   150   155   160

```
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
```

```
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu
785                 790                 795                 800
```

```
Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu
                805                 810                 815

Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile
            820                 825                 830

Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
        835                 840                 845

Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
    850                 855                 860

Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
865                 870                 875                 880

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
                885                 890                 895

Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
            900                 905                 910

His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
        915                 920                 925

Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
    930                 935                 940

Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Glu Glu
945                 950                 955                 960

Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys
                965                 970                 975

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
            980                 985                 990

Gly Leu Ser Glu Leu Asp Lys Ala  Gly Phe Ile Lys Arg  Gln Leu Val
        995                 1000                1005

Glu Thr  Arg Gln Ile Thr Lys  His Val Ala Gln Ile  Leu Asp Ser
    1010                1015                1020
```

```
Arg Met  Asn Thr Lys Tyr Asp  Glu Asn Asp Lys Leu  Ile Arg Glu
    1025                1030                1035

Val Lys  Val Ile Thr Leu Lys  Ser Lys Leu Val Ser  Asp Phe Arg
    1040                1045                1050

Lys Asp  Phe Gln Phe Tyr Lys  Val Arg Glu Ile Asn  Asn Tyr His
    1055                1060                1065

His Ala  His Asp Ala Tyr Leu  Asn Ala Val Val Gly  Thr Ala Leu
    1070                1075                1080

Ile Lys  Lys Tyr Pro Lys Leu  Glu Ser Glu Phe Val  Tyr Gly Asp
    1085                1090                1095

Tyr Lys  Val Tyr Asp Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln
    1100                1105                1110

Glu Ile  Gly Lys Ala Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile
    1115                1120                1125

Met Asn  Phe Phe Lys Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile
    1130                1135                1140

Arg Lys  Arg Pro Leu Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile
    1145                1150                1155

Val Trp  Asp Lys Gly Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu
    1160                1165                1170

Ser Met  Pro Gln Val Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr
    1175                1180                1185

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

Lys Leu  Ile Ala Arg Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly
    1205                1210                1215
```

```
Gly Phe  Asp Ser Pro Thr Val  Ala Tyr Ser Val Leu  Val Val Ala
    1220                1225                1230

Lys Val  Glu Lys Gly Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu
    1235                1240                1245

Leu Leu  Gly Ile Thr Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn
    1250                1255                1260

Pro Ile  Asp Phe Leu Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys
    1265                1270                1275

Asp Leu  Ile Ile Lys Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu
    1280                1285                1290

Asn Gly  Arg Lys Arg Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys
    1295                1300                1305

Gly Asn  Glu Leu Ala Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr
    1310                1315                1320

Leu Ala  Ser His Tyr Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn
    1325                1330                1335

Glu Gln  Lys Gln Leu Phe Val  Glu Gln His Lys His  Tyr Leu Asp
    1340                1345                1350

Glu Ile  Ile Glu Gln Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu
    1355                1360                1365

Ala Asp  Ala Asn Leu Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His
    1370                1375                1380

Arg Asp  Lys Pro Ile Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu
    1385                1390                1395

Phe Thr  Leu Thr Asn Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe
    1400                1405                1410

Asp Thr  Thr Ile Asp Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val
    1415                1420                1425
```

Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu
    1430   1435   1440

Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
    1445   1450

<210> 301
<211> 1420
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 125 Cas9 delta 792-906
       TadAins polypeptide sequence

<400> 301

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

```
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
```

```
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
```

```
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
```

```
Glu Glu Gly Ile Lys Glu Leu Gly Ser Glu Val Glu Phe Ser His Glu
785                 790                 795                 800


Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu
                805                 810                 815


Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile
            820                 825                 830


Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His
        835                 840                 845


Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr
    850                 855                 860


Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met
865                 870                 875                 880


Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly
                885                 890                 895


Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu
            900                 905                 910


His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu
        915                 920                 925


Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg
    930                 935                 940


Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Gly Leu
945                 950                 955                 960


Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr
                965                 970                 975


Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn
            980                 985                 990


Thr Lys Tyr Asp Glu Asn Asp Lys  Leu Ile Arg Glu Val  Lys Val Ile
```

```
        995                 1000                1005

Thr Leu  Lys Ser Lys Leu Val  Ser Asp Phe Arg Lys  Asp Phe Gln
    1010                1015                1020

Phe Tyr  Lys Val Arg Glu Ile  Asn Asn Tyr His His  Ala His Asp
    1025                1030                1035

Ala Tyr  Leu Asn Ala Val Val  Gly Thr Ala Leu Ile  Lys Lys Tyr
    1040                1045                1050

Pro Lys  Leu Glu Ser Glu Phe  Val Tyr Gly Asp Tyr  Lys Val Tyr
    1055                1060                1065

Asp Val  Arg Lys Met Ile Ala  Lys Ser Glu Gln Glu  Ile Gly Lys
    1070                1075                1080

Ala Thr  Ala Lys Tyr Phe Phe  Tyr Ser Asn Ile Met  Asn Phe Phe
    1085                1090                1095

Lys Thr  Glu Ile Thr Leu Ala  Asn Gly Glu Ile Arg  Lys Arg Pro
    1100                1105                1110

Leu Ile  Glu Thr Asn Gly Glu  Thr Gly Glu Ile Val  Trp Asp Lys
    1115                1120                1125

Gly Arg  Asp Phe Ala Thr Val  Arg Lys Val Leu Ser  Met Pro Gln
    1130                1135                1140

Val Asn  Ile Val Lys Lys Thr  Glu Val Gln Thr Gly  Gly Phe Ser
    1145                1150                1155

Lys Glu  Ser Ile Leu Pro Lys  Arg Asn Ser Asp Lys  Leu Ile Ala
    1160                1165                1170

Arg Lys  Lys Asp Trp Asp Pro  Lys Lys Tyr Gly Gly  Phe Asp Ser
    1175                1180                1185

Pro Thr  Val Ala Tyr Ser Val  Leu Val Val Ala Lys  Val Glu Lys
    1190                1195                1200
```

```
Gly Lys  Ser Lys Lys Leu Lys  Ser Val Lys Glu Leu  Leu Gly Ile
    1205                 1210                 1215

Thr Ile  Met Glu Arg Ser Ser  Phe Glu Lys Asn Pro  Ile Asp Phe
    1220                 1225                 1230

Leu Glu  Ala Lys Gly Tyr Lys  Glu Val Lys Lys Asp  Leu Ile Ile
    1235                 1240                 1245

Lys Leu  Pro Lys Tyr Ser Leu  Phe Glu Leu Glu Asn  Gly Arg Lys
    1250                 1255                 1260

Arg Met  Leu Ala Ser Ala Gly  Glu Leu Gln Lys Gly  Asn Glu Leu
    1265                 1270                 1275

Ala Leu  Pro Ser Lys Tyr Val  Asn Phe Leu Tyr Leu  Ala Ser His
    1280                 1285                 1290

Tyr Glu  Lys Leu Lys Gly Ser  Pro Glu Asp Asn Glu  Gln Lys Gln
    1295                 1300                 1305

Leu Phe  Val Glu Gln His Lys  His Tyr Leu Asp Glu  Ile Ile Glu
    1310                 1315                 1320

Gln Ile  Ser Glu Phe Ser Lys  Arg Val Ile Leu Ala  Asp Ala Asn
    1325                 1330                 1335

Leu Asp  Lys Val Leu Ser Ala  Tyr Asn Lys His Arg  Asp Lys Pro
    1340                 1345                 1350

Ile Arg  Glu Gln Ala Glu Asn  Ile Ile His Leu Phe  Thr Leu Thr
    1355                 1360                 1365

Asn Leu  Gly Ala Pro Ala Ala  Phe Lys Tyr Phe Asp  Thr Thr Ile
    1370                 1375                 1380

Asp Arg  Lys Arg Tyr Thr Ser  Thr Lys Glu Val Leu  Asp Ala Thr
    1385                 1390                 1395

Leu Ile  His Gln Ser Ile Thr  Gly Leu Tyr Glu Thr  Arg Ile Asp
```

1400 1405 1410

Leu Ser Gln Leu Gly Gly Asp
1415 1420

<210> 302
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 126 TadA CP65ins 1003 polypeptide sequence

<400> 302

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

```
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
```

```
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575
```

```
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
```

```
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Thr Ala  His Ala Glu
        995                 1000                1005
```

```
Ile Met  Ala Leu Arg Gln Gly  Gly Leu Val Met Gln  Asn Tyr Arg
    1010                1015                1020

Leu Ile  Asp Ala Thr Leu Tyr  Val Thr Phe Glu Pro  Cys Val Met
    1025                1030                1035

Cys Ala  Gly Ala Met Ile His  Ser Arg Ile Gly Arg  Val Val Phe
    1040                1045                1050

Gly Val  Arg Asn Ala Lys Thr  Gly Ala Ala Gly Ser  Leu Met Asp
    1055                1060                1065

Val Leu  His Tyr Pro Gly Met  Asn His Arg Val Glu  Ile Thr Glu
    1070                1075                1080

Gly Ile  Leu Ala Asp Glu Cys  Ala Ala Leu Leu Cys  Tyr Phe Phe
    1085                1090                1095

Arg Met  Pro Arg Gln Val Phe  Asn Ala Gln Lys Lys  Ala Gln Ser
    1100                1105                1110

Ser Thr  Asp Gly Ser Ser Gly  Ser Glu Thr Pro Gly  Thr Ser Glu
    1115                1120                1125

Ser Ala  Thr Pro Glu Ser Ser  Gly Ser Glu Val Glu  Phe Ser His
    1130                1135                1140

Glu Tyr  Trp Met Arg His Ala  Leu Thr Leu Ala Lys  Arg Ala Arg
    1145                1150                1155

Asp Glu  Arg Glu Val Pro Val  Gly Ala Val Leu Val  Leu Asn Asn
    1160                1165                1170

Arg Val  Ile Gly Glu Gly Trp  Asn Arg Ala Ile Gly  Leu His Asp
    1175                1180                1185

Pro Leu  Glu Ser Glu Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp
    1190                1195                1200

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215
```

```
Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410
```

```
Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425


Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440


Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455


Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470


Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485


Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500


Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515


Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530


Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545


Ser Gln  Leu Gly Gly Asp
    1550


<210>  303
<211>  1554
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 127 TadA CP65ins 1016
       polypeptide sequence

<400>  303

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15
```

```
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
```

```
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445
```

```
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
```

```
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
```

```
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
               885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
               965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Thr Ala His Ala Glu  Ile Met Ala
    1010                1015                1020

Leu Arg  Gln Gly Gly Leu Val  Met Gln Asn Tyr Arg  Leu Ile Asp
    1025                1030                1035

Ala Thr  Leu Tyr Val Thr Phe  Glu Pro Cys Val Met  Cys Ala Gly
    1040                1045                1050

Ala Met  Ile His Ser Arg Ile  Gly Arg Val Val Phe  Gly Val Arg
    1055                1060                1065

Asn Ala  Lys Thr Gly Ala Ala  Gly Ser Leu Met Asp  Val Leu His
    1070                1075                1080

Tyr Pro  Gly Met Asn His Arg  Val Glu Ile Thr Glu  Gly Ile Leu
```

```
    1085                1090                1095

Ala Asp  Glu Cys Ala Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro
    1100                1105                1110

Arg Gln  Val Phe Asn Ala Gln  Lys Lys Ala Gln Ser  Ser Thr Asp
    1115                1120                1125

Gly Ser  Ser Gly Ser Glu Thr  Pro Gly Thr Ser Glu  Ser Ala Thr
    1130                1135                1140

Pro Glu  Ser Ser Gly Ser Glu  Val Glu Phe Ser His  Glu Tyr Trp
    1145                1150                1155

Met Arg  His Ala Leu Thr Leu  Ala Lys Arg Ala Arg  Asp Glu Arg
    1160                1165                1170

Glu Val  Pro Val Gly Ala Val  Leu Val Leu Asn Asn  Arg Val Ile
    1175                1180                1185

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

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290
```

```
Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                 1300                 1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                 1315                 1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                 1330                 1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                 1345                 1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                 1360                 1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                 1375                 1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                 1390                 1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                 1405                 1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                 1420                 1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                 1435                 1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                 1450                 1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                 1465                 1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                 1480                 1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
```

1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 304
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 128 TadA CP65ins 1022 polypeptide sequence

<400> 304

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

```
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
```

```
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
```

```
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
```

```
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
```

```
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
               965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
           980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Thr
    1010                1015                1020

Ala His  Ala Glu Ile Met Ala  Leu Arg Gln Gly Gly  Leu Val Met
    1025                1030                1035

Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu Tyr Val  Thr Phe Glu
    1040                1045                1050

Pro Cys  Val Met Cys Ala Gly  Ala Met Ile His Ser  Arg Ile Gly
    1055                1060                1065

Arg Val  Val Phe Gly Val Arg  Asn Ala Lys Thr Gly  Ala Ala Gly
    1070                1075                1080

Ser Leu  Met Asp Val Leu His  Tyr Pro Gly Met Asn  His Arg Val
    1085                1090                1095

Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu Cys Ala  Ala Leu Leu
    1100                1105                1110

Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val Phe Asn  Ala Gln Lys
    1115                1120                1125

Lys Ala  Gln Ser Ser Thr Asp  Gly Ser Ser Gly Ser  Glu Thr Pro
    1130                1135                1140

Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser Ser Gly  Ser Glu Val
    1145                1150                1155

Glu Phe  Ser His Glu Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala
    1160                1165                1170
```

```
Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro Val Gly  Ala Val Leu
    1175                1180                1185

Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile
    1190                1195                1200

Gly Leu  His Asp Pro Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365
```

```
Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545

Ser Gln  Leu Gly Gly Asp
    1550
```

<210> 305
<211> 1554

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 129 TadA CP65ins 1029
polypeptide sequence

<400> 305

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

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

```
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400
```

```
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
```

```
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
```

```
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990

Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Thr  Ala His Ala Glu Ile  Met Ala Leu
    1025                1030                1035

Arg Gln  Gly Gly Leu Val Met  Gln Asn Tyr Arg Leu  Ile Asp Ala
```

```
    1040                1045                1050

Thr Leu  Tyr Val Thr Phe Glu  Pro Cys Val Met Cys  Ala Gly Ala
    1055                1060                1065

Met Ile  His Ser Arg Ile Gly  Arg Val Val Phe Gly  Val Arg Asn
    1070                1075                1080

Ala Lys  Thr Gly Ala Ala Gly  Ser Leu Met Asp Val  Leu His Tyr
    1085                1090                1095

Pro Gly  Met Asn His Arg Val  Glu Ile Thr Glu Gly  Ile Leu Ala
    1100                1105                1110

Asp Glu  Cys Ala Ala Leu Leu  Cys Tyr Phe Phe Arg  Met Pro Arg
    1115                1120                1125

Gln Val  Phe Asn Ala Gln Lys  Lys Ala Gln Ser Ser  Thr Asp Gly
    1130                1135                1140

Ser Ser  Gly Ser Glu Thr Pro  Gly Thr Ser Glu Ser  Ala Thr Pro
    1145                1150                1155

Glu Ser  Ser Gly Ser Glu Val  Glu Phe Ser His Glu  Tyr Trp Met
    1160                1165                1170

Arg His  Ala Leu Thr Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu
    1175                1180                1185

Val Pro  Val Gly Ala Val Leu  Val Leu Asn Asn Arg  Val Ile Gly
    1190                1195                1200

Glu Gly  Trp Asn Arg Ala Ile  Gly Leu His Asp Pro  Gly Lys Ala
    1205                1210                1215

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245
```

```
Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
```

```
    1445                1450                1455


Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470


Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485


Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500


Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515


Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530


Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1535                1540                1545


Ser Gln  Leu Gly Gly Asp
    1550


<210>  306
<211>  1554
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: 130 TadA CP65ins 1041
       polypeptide sequence

<400>  306

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15


Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30


Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
```

```
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
```

```
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480
```

```
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
```

```
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
```

```
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925


Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940


Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Thr Ala His Ala Glu  Ile Met Ala Leu Arg  Gln Gly Gly
    1040                1045                1050


Leu Val  Met Gln Asn Tyr Arg  Leu Ile Asp Ala Thr  Leu Tyr Val
    1055                1060                1065


Thr Phe  Glu Pro Cys Val Met  Cys Ala Gly Ala Met  Ile His Ser
    1070                1075                1080


Arg Ile  Gly Arg Val Val Phe  Gly Val Arg Asn Ala  Lys Thr Gly
    1085                1090                1095


Ala Ala  Gly Ser Leu Met Asp  Val Leu His Tyr Pro  Gly Met Asn
    1100                1105                1110


His Arg  Val Glu Ile Thr Glu  Gly Ile Leu Ala Asp  Glu Cys Ala
    1115                1120                1125
```

```
Ala Leu  Leu Cys Tyr Phe Phe  Arg Met Pro Arg Gln  Val Phe Asn
    1130                1135                1140

Ala Gln  Lys Lys Ala Gln Ser  Ser Thr Asp Gly Ser  Ser Gly Ser
    1145                1150                1155

Glu Thr  Pro Gly Thr Ser Glu  Ser Ala Thr Pro Glu  Ser Ser Gly
    1160                1165                1170

Ser Glu  Val Glu Phe Ser His  Glu Tyr Trp Met Arg  His Ala Leu
    1175                1180                1185

Thr Leu  Ala Lys Arg Ala Arg  Asp Glu Arg Glu Val  Pro Val Gly
    1190                1195                1200

Ala Val  Leu Val Leu Asn Asn  Arg Val Ile Gly Glu  Gly Trp Asn
    1205                1210                1215

Arg Ala  Ile Gly Leu His Asp  Pro Asn Ile Met Asn  Phe Phe Lys
    1220                1225                1230

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                1240                1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                1255                1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                1270                1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                1285                1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                1300                1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                1315                1320
```

```
Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                1330                1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                1345                1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                1360                1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                1375                1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                1390                1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1400                1405                1410

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1415                1420                1425

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1490                1495                1500

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1505                1510                1515

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1520                1525                1530
```

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 307
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 131 TadA CP65ins 1054
       polypeptide sequence

<400> 307

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1   5   10   15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp

```
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
```

```
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
```

```
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
```

```
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800


Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815


Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830


Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845


Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860


Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880


Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895


Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910


Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925


Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940


Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
```

```
        995                 1000                1005

Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

Asn Thr  Ala His Ala Glu Ile  Met Ala Leu Arg Gln  Gly Gly Leu
    1055                1060                1065

Val Met  Gln Asn Tyr Arg Leu  Ile Asp Ala Thr Leu  Tyr Val Thr
    1070                1075                1080

Phe Glu  Pro Cys Val Met Cys  Ala Gly Ala Met Ile  His Ser Arg
    1085                1090                1095

Ile Gly  Arg Val Val Phe Gly  Val Arg Asn Ala Lys  Thr Gly Ala
    1100                1105                1110

Ala Gly  Ser Leu Met Asp Val  Leu His Tyr Pro Gly  Met Asn His
    1115                1120                1125

Arg Val  Glu Ile Thr Glu Gly  Ile Leu Ala Asp Glu  Cys Ala Ala
    1130                1135                1140

Leu Leu  Cys Tyr Phe Phe Arg  Met Pro Arg Gln Val  Phe Asn Ala
    1145                1150                1155

Gln Lys  Lys Ala Gln Ser Ser  Thr Asp Gly Ser Ser  Gly Ser Glu
    1160                1165                1170

Thr Pro  Gly Thr Ser Glu Ser  Ala Thr Pro Glu Ser  Ser Gly Ser
    1175                1180                1185

Glu Val  Glu Phe Ser His Glu  Tyr Trp Met Arg His  Ala Leu Thr
    1190                1195                1200
```

```
Leu Ala  Lys Arg Ala Arg Asp  Glu Arg Glu Val Pro  Val Gly Ala
    1205                 1210                 1215

Val Leu  Val Leu Asn Asn Arg  Val Ile Gly Glu Gly  Trp Asn Arg
    1220                 1225                 1230

Ala Ile  Gly Leu His Asp Pro  Gly Glu Ile Arg Lys  Arg Pro Leu
    1235                 1240                 1245

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1250                 1255                 1260

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1265                 1270                 1275

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1280                 1285                 1290

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1295                 1300                 1305

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1310                 1315                 1320

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1325                 1330                 1335

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1340                 1345                 1350

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1355                 1360                 1365

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1370                 1375                 1380

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1385                 1390                 1395

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
```

1400 1405 1410

Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1415 1420 1425

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1430 1435 1440

Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1445 1450 1455

Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1460 1465 1470

Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1475 1480 1485

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1490 1495 1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1505 1510 1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1520 1525 1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1535 1540 1545

Ser Gln Leu Gly Gly Asp
1550

<210> 308
<211> 1554
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 132 TadA CP65ins 1246 polypeptide sequence

<400> 308

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                   10                  15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220
```

```
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430
```

```
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655
```

```
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
    850                 855                 860
```

```
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880


Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895


Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910


Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925


Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940


Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960


Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975


Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990


Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005


Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020


Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035


Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050


Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065


Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080
```

```
Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

Arg Asn  Ser Asp Lys Leu Ile  Ala Arg Lys Lys Asp  Trp Asp Pro
    1115                1120                1125

Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

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

Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200

Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Thr
    1235                1240                1245

Ala His  Ala Glu Ile Met Ala  Leu Arg Gln Gly Gly  Leu Val Met
    1250                1255                1260

Gln Asn  Tyr Arg Leu Ile Asp  Ala Thr Leu Tyr Val  Thr Phe Glu
    1265                1270                1275
```

```
Pro Cys  Val Met Cys Ala Gly  Ala Met Ile His Ser  Arg Ile Gly
    1280                1285                1290

Arg Val  Val Phe Gly Val Arg  Asn Ala Lys Thr Gly  Ala Ala Gly
    1295                1300                1305

Ser Leu  Met Asp Val Leu His  Tyr Pro Gly Met Asn  His Arg Val
    1310                1315                1320

Glu Ile  Thr Glu Gly Ile Leu  Ala Asp Glu Cys Ala  Ala Leu Leu
    1325                1330                1335

Cys Tyr  Phe Phe Arg Met Pro  Arg Gln Val Phe Asn  Ala Gln Lys
    1340                1345                1350

Lys Ala  Gln Ser Ser Thr Asp  Gly Ser Ser Gly Ser  Glu Thr Pro
    1355                1360                1365

Gly Thr  Ser Glu Ser Ala Thr  Pro Glu Ser Ser Gly  Ser Glu Val
    1370                1375                1380

Glu Phe  Ser His Glu Tyr Trp  Met Arg His Ala Leu  Thr Leu Ala
    1385                1390                1395

Lys Arg  Ala Arg Asp Glu Arg  Glu Val Pro Val Gly  Ala Val Leu
    1400                1405                1410

Val Leu  Asn Asn Arg Val Ile  Gly Glu Gly Trp Asn  Arg Ala Ile
    1415                1420                1425

Gly Leu  His Asp Pro Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1430                1435                1440

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1445                1450                1455

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1460                1465                1470

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1475                1480                1485
```

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
    1490   1495   1500

Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
    1505   1510   1515

Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
    1520   1525   1530

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
    1535   1540   1545

Ser Gln Leu Gly Gly Asp
    1550

<210> 309
<211> 160
<212> PRT
<213> Staphylococcus aureus

<220>
<221> MISC_FEATURE
<222> (1)..(160)
<223> TadA polypeptide sequence

<400> 309

Met Gly Ser His Met Thr Asn Asp Ile Tyr Phe Met Thr Leu Ala Ile
1   5   10   15

Glu Glu Ala Lys Lys Ala Ala Gln Leu Gly Glu Val Pro Ile Gly Ala
            20   25   30

Ile Ile Thr Lys Asp Asp Glu Val Ile Ala Arg Ala His Asn Leu Arg
        35   40   45

Glu Thr Leu Gln Gln Pro Thr Ala His Ala Glu His Ile Ala Ile Glu
    50   55   60

Arg Ala Ala Lys Val Leu Gly Ser Trp Arg Leu Glu Gly Cys Thr Leu
65   70   75   80

```
Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Thr Ile Val Met
                85                  90                  95

Ser Arg Ile Pro Arg Val Val Tyr Gly Ala Asp Asp Pro Lys Gly Gly
            100                 105                 110

Cys Ser Gly Ser Leu Met Asn Leu Leu Gln Gln Ser Asn Phe Asn His
        115                 120                 125

Arg Ala Ile Val Asp Lys Gly Val Leu Lys Glu Ala Cys Ser Thr Leu
    130                 135                 140

Leu Thr Thr Phe Phe Lys Asn Leu Arg Ala Asn Lys Lys Ser Thr Asn
145                 150                 155                 160
```

<210> 310
<211> 161
<212> PRT
<213> Bacillus subtilis

<220>
<221> MISC_FEATURE
<222> (1)..(161)
<223> TadA polypeptide sequence

<400> 310

```
Met Thr Gln Asp Glu Leu Tyr Met Lys Glu Ala Ile Lys Glu Ala Lys
1               5                   10                  15

Lys Ala Glu Glu Lys Gly Glu Val Pro Ile Gly Ala Val Leu Val Ile
            20                  25                  30

Asn Gly Glu Ile Ile Ala Arg Ala His Asn Leu Arg Glu Thr Glu Gln
        35                  40                  45

Arg Ser Ile Ala His Ala Glu Met Leu Val Ile Asp Glu Ala Cys Lys
    50                  55                  60

Ala Leu Gly Thr Trp Arg Leu Glu Gly Ala Thr Leu Tyr Val Thr Leu
65                  70                  75                  80
```

```
Glu Pro Cys Pro Met Cys Ala Gly Ala Val Val Leu Ser Arg Val Glu
                85                  90                  95

Lys Val Val Phe Gly Ala Phe Asp Pro Lys Gly Gly Cys Ser Gly Thr
            100                 105                 110

Leu Met Asn Leu Leu Gln Glu Glu Arg Phe Asn His Gln Ala Glu Val
        115                 120                 125

Val Ser Gly Val Leu Glu Glu Glu Cys Gly Gly Met Leu Ser Ala Phe
    130                 135                 140

Phe Arg Glu Leu Arg Lys Lys Lys Lys Ala Ala Arg Lys Asn Leu Ser
145                 150                 155                 160

Glu
```

<210> 311
<211> 183
<212> PRT
<213> Salmonella typhimurium

<220>
<221> MISC_FEATURE
<222> (1)..(183)
<223> TadA polypeptide sequence

<400> 311

```
Met Pro Pro Ala Phe Ile Thr Gly Val Thr Ser Leu Ser Asp Val Glu
1               5                   10                  15

Leu Asp His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
            20                  25                  30

Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn
        35                  40                  45

His Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp
    50                  55                  60

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
```

65 70 75 80

Leu Gln Asn Tyr Arg Leu Leu Asp Thr Thr Leu Tyr Val Thr Leu Glu
85 90 95

Pro Cys Val Met Cys Ala Gly Ala Met Val His Ser Arg Ile Gly Arg
100 105 110

Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu
115 120 125

Ile Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Ile
130 135 140

Glu Gly Val Leu Arg Asp Glu Cys Ala Thr Leu Leu Ser Asp Phe Phe
145 150 155 160

Arg Met Arg Arg Gln Glu Ile Lys Ala Leu Lys Lys Ala Asp Arg Ala
165 170 175

Glu Gly Ala Gly Pro Ala Val
180

<210> 312
<211> 164
<212> PRT
<213> Shewanella putrefaciens

<220>
<221> MISC_FEATURE
<222> (1)..(164)
<223> TadA polypeptide sequence

<400> 312

Met Asp Glu Tyr Trp Met Gln Val Ala Met Gln Met Ala Glu Lys Ala
1 5 10 15

Glu Ala Ala Gly Glu Val Pro Val Gly Ala Val Leu Val Lys Asp Gly
20 25 30

Gln Gln Ile Ala Thr Gly Tyr Asn Leu Ser Ile Ser Gln His Asp Pro
35 40 45

```
Thr Ala His Ala Glu Ile Leu Cys Leu Arg Ser Ala Gly Lys Lys Leu
    50                  55                  60

Glu Asn Tyr Arg Leu Leu Asp Ala Thr Leu Tyr Ile Thr Leu Glu Pro
65                  70                  75                  80

Cys Ala Met Cys Ala Gly Ala Met Val His Ser Arg Ile Ala Arg Val
                85                  90                  95

Val Tyr Gly Ala Arg Asp Glu Lys Thr Gly Ala Ala Gly Thr Val Val
            100                 105                 110

Asn Leu Leu Gln His Pro Ala Phe Asn His Gln Val Glu Val Thr Ser
        115                 120                 125

Gly Val Leu Ala Glu Ala Cys Ser Ala Gln Leu Ser Arg Phe Phe Lys
    130                 135                 140

Arg Arg Arg Asp Glu Lys Lys Ala Leu Lys Leu Ala Gln Arg Ala Gln
145                 150                 155                 160

Gln Gly Ile Glu
```

<210> 313
<211> 173
<212> PRT
<213> Haemophilus influenzae

<220>
<221> MISC_FEATURE
<222> (1)..(173)
<223> TadA polypeptide sequence

<400> 313

```
Met Asp Ala Ala Lys Val Arg Ser Glu Phe Asp Glu Lys Met Met Arg
1               5                   10                  15

Tyr Ala Leu Glu Leu Ala Asp Lys Ala Glu Ala Leu Gly Glu Ile Pro
            20                  25                  30
```

```
Val Gly Ala Val Leu Val Asp Asp Ala Arg Asn Ile Ile Gly Glu Gly
        35                  40                  45

Trp Asn Leu Ser Ile Val Gln Ser Asp Pro Thr Ala His Ala Glu Ile
    50                  55                  60

Ile Ala Leu Arg Asn Gly Ala Lys Asn Ile Gln Asn Tyr Arg Leu Leu
65                  70                  75                  80

Asn Ser Thr Leu Tyr Val Thr Leu Glu Pro Cys Thr Met Cys Ala Gly
                85                  90                  95

Ala Ile Leu His Ser Arg Ile Lys Arg Leu Val Phe Gly Ala Ser Asp
            100                 105                 110

Tyr Lys Thr Gly Ala Ile Gly Ser Arg Phe His Phe Phe Asp Asp Tyr
        115                 120                 125

Lys Met Asn His Thr Leu Glu Ile Thr Ser Gly Val Leu Ala Glu Glu
    130                 135                 140

Cys Ser Gln Lys Leu Ser Thr Phe Phe Gln Lys Arg Arg Glu Glu Lys
145                 150                 155                 160

Lys Ile Glu Lys Ala Leu Leu Lys Ser Leu Ser Asp Lys
                165                 170
```

<210> 314
<211> 161
<212> PRT
<213> Caulobacter crescentus

<220>
<221> MISC_FEATURE
<222> (1)..(161)
<223> TadA polypeptide sequence

<400> 314

```
Met Arg Thr Asp Glu Ser Glu Asp Gln Asp His Arg Met Met Arg Leu
1               5                   10                  15
```

Ala Leu Asp Ala Ala Arg Ala Ala Ala Glu Ala Gly Glu Thr Pro Val
            20   25   30

Gly Ala Val Ile Leu Asp Pro Ser Thr Gly Glu Val Ile Ala Thr Ala
        35   40   45

Gly Asn Gly Pro Ile Ala Ala His Asp Pro Thr Ala His Ala Glu Ile
    50   55   60

Ala Ala Met Arg Ala Ala Ala Ala Lys Leu Gly Asn Tyr Arg Leu Thr
65   70   75   80

Asp Leu Thr Leu Val Val Thr Leu Glu Pro Cys Ala Met Cys Ala Gly
                85   90   95

Ala Ile Ser His Ala Arg Ile Gly Arg Val Val Phe Gly Ala Asp Asp
            100   105   110

Pro Lys Gly Gly Ala Val Val His Gly Pro Lys Phe Phe Ala Gln Pro
        115   120   125

Thr Cys His Trp Arg Pro Glu Val Thr Gly Gly Val Leu Ala Asp Glu
    130   135   140

Ser Ala Asp Leu Leu Arg Gly Phe Phe Arg Ala Arg Arg Lys Ala Lys
145   150   155   160

Ile

<210> 315
<211> 179
<212> PRT
<213> Geobacter sulfurreducens

<220>
<221> MISC_FEATURE
<222> (1)..(179)
<223> TadA polypeptide sequence

<400> 315

Met Ser Ser Leu Lys Lys Thr Pro Ile Arg Asp Asp Ala Tyr Trp Met

```
1               5                   10                  15

Gly Lys Ala Ile Arg Glu Ala Ala Lys Ala Ala Ala Arg Asp Glu Val
            20                  25                  30

Pro Ile Gly Ala Val Ile Val Arg Asp Gly Ala Val Ile Gly Arg Gly
        35                  40                  45

His Asn Leu Arg Glu Gly Ser Asn Asp Pro Ser Ala His Ala Glu Met
    50                  55                  60

Ile Ala Ile Arg Gln Ala Ala Arg Arg Ser Ala Asn Trp Arg Leu Thr
65                  70                  75                  80

Gly Ala Thr Leu Tyr Val Thr Leu Glu Pro Cys Leu Met Cys Met Gly
                85                  90                  95

Ala Ile Ile Leu Ala Arg Leu Glu Arg Val Val Phe Gly Cys Tyr Asp
            100                 105                 110

Pro Lys Gly Gly Ala Ala Gly Ser Leu Tyr Asp Leu Ser Ala Asp Pro
        115                 120                 125

Arg Leu Asn His Gln Val Arg Leu Ser Pro Gly Val Cys Gln Glu Glu
    130                 135                 140

Cys Gly Thr Met Leu Ser Asp Phe Phe Arg Asp Leu Arg Arg Arg Lys
145                 150                 155                 160

Lys Ala Lys Ala Thr Pro Ala Leu Phe Ile Asp Glu Arg Lys Val Pro
                165                 170                 175

Pro Glu Pro
```

<210> 316
<211> 167
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: TadA*8 polypeptide sequence

<400> 316

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 Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp
165

<210> 317
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: gRNA scaffold polynucleotide

sequence

<400> 317
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 318
<211> 125
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: gRNA scaffold polynucleotide sequence

<400> 318
guuuuuguac ucucaagauu uaaguaacug uacaacgaaa cuuacacagu uacuuaaauc 60

uugcagaagc uacaaagaua aggcuucaug ccgaaaucaa cacccuguca uuuuauggca 120

gggug 125

<210> 319
<211> 76
<212> RNA
<213> S. pyogenes

<220>
<221> misc_feature
<222> (1)..(76)
<223> S. pyogenes gRNA scaffold ologonucleotide sequence

<400> 319
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugc 76

<210> 320
<211> 84
<212> RNA
<213> S. aureus

<220>
<221> misc_feature
<222> (1)..(84)
<223> S. aureus gRNA scaffold oligonucleotide sequence

<400> 320

guuuuaguac ucuguaauga aaauuacaga aucuacuaaa acaaggcaaa augccguguu 60

uaucucguca acuuguuggc gaga 84

<210> 321
<211> 97
<212> RNA
<213> Bacillus hisashii

<220>
<221> misc_feature
<222> (1)..(97)
<223> BhCas12b gRNA scaffold oligonucleotide sequence

<400> 321
guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcac 97

<210> 322
<211> 88
<212> RNA
<213> Bacillus sp.

<220>
<221> misc_feature
<222> (1)..(88)
<223> BvCas12b gRNA scaffold oligonucleotide sequence

<400> 322
gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcac 88

<210> 323
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: gRNA scaffold nucleotide sequence; oligonucleotide sequence

<220>
<221> modified_base
<222> (78)..(80)
<223> um

<220>
<221> misc_feature
<222> (78)..(80)
<223> In embodiments, a nucleotide(s) is linked to the preceding 5' nucleotide by a 3' phosphorothioate internucleotide liknage

<400> 323
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

gggaccgagu cggugcuuuu 80

<210> 324
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: gRNA scaffold nucleotide sequence

<220>
<221> misc_feature
<222> (78)..(80)
<223> In embodiments, a nucleotide(s) is linked to the preceding 5' nucleotide by a 3' phosphorothioate internucleotide liknage

<220>
<221> modified_base
<222> (78)..(80)
<223> um

<400> 324
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 325
<211> 117
<212> RNA
<213> Bacillus hisashii

<220>
<221> misc_feature
<222> (1)..(117)
<223> BhCas12b gRNA scaffold + guide sequence; oligonucleotide sequence

<220>
<221> misc_feature
<222> (1)..(97)
<223> BhCas12b sgRNA scaffold

<400> 325
guucugucuu uuggucagga caaccgucua gcuauaagug cugcagggug ugagaaacuc 60

cuauugcugg acgaugucuc uuacgaggca uuagcacnnn nnnnnnnnnn nnnnnnn 117

<210> 326
<211> 108
<212> RNA
<213> Bacillus sp.

<220>
<221> misc_feature
<222> (1)..(108)
<223> BvCas12b gRNA scaffold + guide sequence; oligonucleotide sequence

<220>
<221> misc_feature
<222> (1)..(88)
<223> BvCas12b gRNA scaffold

<400> 326
gaccuauagg gucaaugaau cugugcgugu gccauaagua auuaaaaauu acccaccaca 60

ggagcaccug aaaacaggug cuuggcacnn nnnnnnnnnn nnnnnnnn 108

<210> 327
<211> 119
<212> RNA
<213> Alicyclobacillus acidiphilus

<220>
<221> misc_feature
<222> (1)..(119)
<223> AaCas12b gRNA scaffold + guide sequence; oligonucleotide sequence

<220>
<221> misc_feature
<222> (1)..(99)
<223> AaCas12b gRNA scaffold

<400> 327
gucuaaagga cagaauuuuu caacgggugu gccaauggcc acuuuccagg uggcaaagcc 60

cguugaacuu cucaaaaaga acgaucugag aaguggcacn nnnnnnnnnn nnnnnnnnn 119

<210> 328
<211> 29
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 328

Pro Lys Lys Lys Arg Lys Val Glu Gly Ala Asp Lys Arg Thr Ala Asp
1 5 10 15

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
20 25

<210> 329
<211> 23
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 329

Arg Lys Ser Gly Lys Ile Ala Ala Ile Val Val Lys Arg Pro Arg Lys
1 5 10 15

Pro Lys Lys Lys Arg Lys Val
20

<210> 330
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 330

Ser Gly Gly Ser Ser Gly Gly Ser
1 5

<210> 331
<211> 1609
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of artificial sequence: pNMG-B335
ABE8.1_Y147T_CP5_NGC PAM_monomer polypeptide sequence

<400> 331

```
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 Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Glu Ile Gly Lys Ala Thr Ala Lys Tyr
```

```
        195                 200                 205

Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu
    210                 215                 220

Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
225                 230                 235                 240

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
                245                 250                 255

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val
            260                 265                 270

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

Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly
    290                 295                 300

Gly Phe Met Gln Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys
305                 310                 315                 320

Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu
                325                 330                 335

Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp
            340                 345                 350

Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile
        355                 360                 365

Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
    370                 375                 380

Met Leu Ala Ser Ala Lys Phe Leu Gln Lys Gly Asn Glu Leu Ala Leu
385                 390                 395                 400

Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
                405                 410                 415
```

```
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
            420                 425                 430

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe
        435                 440                 445

Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser
    450                 455                 460

Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
465                 470                 475                 480

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Arg Ala Phe
                485                 490                 495

Lys Tyr Phe Asp Thr Thr Ile Ala Arg Lys Glu Tyr Arg Ser Thr Lys
            500                 505                 510

Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr
        515                 520                 525

Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Gly Gly Ser Gly
    530                 535                 540

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
545                 550                 555                 560

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
                565                 570                 575

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            580                 585                 590

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        595                 600                 605

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    610                 615                 620

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
```

625 630 635 640

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
645 650 655

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
660 665 670

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
675 680 685

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
690 695 700

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
705 710 715 720

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
725 730 735

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
740 745 750

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
755 760 765

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
770 775 780

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
785 790 795 800

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
805 810 815

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
820 825 830

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
835 840 845

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

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
865                 870                 875                 880

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                885                 890                 895

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            900                 905                 910

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        915                 920                 925

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    930                 935                 940

Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
945                 950                 955                 960

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                965                 970                 975

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            980                 985                 990

Leu Lys Asp Asn Arg Glu Lys Ile  Glu Lys Ile Leu Thr  Phe Arg Ile
        995                 1000                1005

Pro Tyr  Tyr Val Gly Pro Leu  Ala Arg Gly Asn Ser  Arg Phe Ala
    1010                1015                1020

Trp Met  Thr Arg Lys Ser Glu  Glu Thr Ile Thr Pro  Trp Asn Phe
    1025                1030                1035

Glu Glu  Val Val Asp Lys Gly  Ala Ser Ala Gln Ser  Phe Ile Glu
    1040                1045                1050

Arg Met  Thr Asn Phe Asp Lys  Asn Leu Pro Asn Glu  Lys Val Leu
```

```
    1055                1060                1065

Pro Lys  His Ser Leu Leu Tyr  Glu Tyr Phe Thr Val  Tyr Asn Glu
    1070                1075                1080

Leu Thr  Lys Val Lys Tyr Val  Thr Glu Gly Met Arg  Lys Pro Ala
    1085                1090                1095

Phe Leu  Ser Gly Glu Gln Lys  Lys Ala Ile Val Asp  Leu Leu Phe
    1100                1105                1110

Lys Thr  Asn Arg Lys Val Thr  Val Lys Gln Leu Lys  Glu Asp Tyr
    1115                1120                1125

Phe Lys  Lys Ile Glu Cys Phe  Asp Ser Val Glu Ile  Ser Gly Val
    1130                1135                1140

Glu Asp  Arg Phe Asn Ala Ser  Leu Gly Thr Tyr His  Asp Leu Leu
    1145                1150                1155

Lys Ile  Ile Lys Asp Lys Asp  Phe Leu Asp Asn Glu  Glu Asn Glu
    1160                1165                1170

Asp Ile  Leu Glu Asp Ile Val  Leu Thr Leu Thr Leu  Phe Glu Asp
    1175                1180                1185

Arg Glu  Met Ile Glu Glu Arg  Leu Lys Thr Tyr Ala  His Leu Phe
    1190                1195                1200

Asp Asp  Lys Val Met Lys Gln  Leu Lys Arg Arg Arg  Tyr Thr Gly
    1205                1210                1215

Trp Gly  Arg Leu Ser Arg Lys  Leu Ile Asn Gly Ile  Arg Asp Lys
    1220                1225                1230

Gln Ser  Gly Lys Thr Ile Leu  Asp Phe Leu Lys Ser  Asp Gly Phe
    1235                1240                1245

Ala Asn  Arg Asn Phe Met Gln  Leu Ile His Asp Asp  Ser Leu Thr
    1250                1255                1260
```

```
Phe Lys  Glu Asp Ile Gln Lys  Ala Gln Val Ser Gly  Gln Gly Asp
    1265                1270                1275

Ser Leu  His Glu His Ile Ala  Asn Leu Ala Gly Ser  Pro Ala Ile
    1280                1285                1290

Lys Lys  Gly Ile Leu Gln Thr  Val Lys Val Val Asp  Glu Leu Val
    1295                1300                1305

Lys Val  Met Gly Arg His Lys  Pro Glu Asn Ile Val  Ile Glu Met
    1310                1315                1320

Ala Arg  Glu Asn Gln Thr Thr  Gln Lys Gly Gln Lys  Asn Ser Arg
    1325                1330                1335

Glu Arg  Met Lys Arg Ile Glu  Glu Gly Ile Lys Glu  Leu Gly Ser
    1340                1345                1350

Gln Ile  Leu Lys Glu His Pro  Val Glu Asn Thr Gln  Leu Gln Asn
    1355                1360                1365

Glu Lys  Leu Tyr Leu Tyr Tyr  Leu Gln Asn Gly Arg  Asp Met Tyr
    1370                1375                1380

Val Asp  Gln Glu Leu Asp Ile  Asn Arg Leu Ser Asp  Tyr Asp Val
    1385                1390                1395

Asp His  Ile Val Pro Gln Ser  Phe Leu Lys Asp Asp  Ser Ile Asp
    1400                1405                1410

Asn Lys  Val Leu Thr Arg Ser  Asp Lys Asn Arg Gly  Lys Ser Asp
    1415                1420                1425

Asn Val  Pro Ser Glu Glu Val  Val Lys Lys Met Lys  Asn Tyr Trp
    1430                1435                1440

Arg Gln  Leu Leu Asn Ala Lys  Leu Ile Thr Gln Arg  Lys Phe Asp
    1445                1450                1455

Asn Leu  Thr Lys Ala Glu Arg  Gly Gly Leu Ser Glu  Leu Asp Lys
```

```
    1460                1465                1470

Ala Gly  Phe Ile Lys Arg Gln  Leu Val Glu Thr Arg  Gln Ile Thr
    1475                1480                1485

Lys His  Val Ala Gln Ile Leu  Asp Ser Arg Met Asn  Thr Lys Tyr
    1490                1495                1500

Asp Glu  Asn Asp Lys Leu Ile  Arg Glu Val Lys Val  Ile Thr Leu
    1505                1510                1515

Lys Ser  Lys Leu Val Ser Asp  Phe Arg Lys Asp Phe  Gln Phe Tyr
    1520                1525                1530

Lys Val  Arg Glu Ile Asn Asn  Tyr His His Ala His  Asp Ala Tyr
    1535                1540                1545

Leu Asn  Ala Val Val Gly Thr  Ala Leu Ile Lys Lys  Tyr Pro Lys
    1550                1555                1560

Leu Glu  Ser Glu Phe Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val
    1565                1570                1575

Arg Lys  Met Ile Ala Lys Ser  Glu Gln Glu Gly Ala  Asp Lys Arg
    1580                1585                1590

Thr Ala  Asp Gly Ser Glu Phe  Glu Ser Pro Lys Lys  Lys Arg Lys
    1595                1600                1605

Val
```

<210> 332
<211> 1807
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: pNMG-357_ABE8.14 with NGC PAM CP5 polypeptide sequence

<220>

<221> MISC_FEATURE
<222> (1)..(174)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (175)..(198)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (199)..(365)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (366)..(397)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (398)..(738)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (739)..(759)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (760)..(1785)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1786)..(1807)
<223> Bipartite nuclear localization sequence

<400> 332

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 Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg 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 Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp 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 Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp 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 Met Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270
```

```
Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
    290                 295                 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Thr Phe Phe Arg Met Pro Arg Gln
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Glu Ile Gly
385                 390                 395                 400

Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe
                405                 410                 415

Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu
            420                 425                 430

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg
        435                 440                 445

Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile
    450                 455                 460

Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
465                 470                 475                 480

Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp
```

```
               485                 490                 495

Asp Pro Lys Lys Tyr Gly Gly Phe Met Gln Pro Thr Val Ala Tyr Ser
            500                 505                 510

Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
        515                 520                 525

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
    530                 535                 540

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

Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu
                565                 570                 575

Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Lys Phe Leu Gln Lys
            580                 585                 590

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

Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln
    610                 615                 620

Lys Gln Leu Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile
625                 630                 635                 640

Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn
                645                 650                 655

Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
            660                 665                 670

Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu
        675                 680                 685

Gly Ala Pro Arg Ala Phe Lys Tyr Phe Asp Thr Thr Ile Ala Arg Lys
    690                 695                 700
```

```
Glu Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
705                 710                 715                 720


Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly
                725                 730                 735


Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly
            740                 745                 750


Ser Gly Gly Ser Gly Gly Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        755                 760                 765


Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    770                 775                 780


Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
785                 790                 795                 800


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


Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            820                 825                 830


Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        835                 840                 845


Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    850                 855                 860


Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
865                 870                 875                 880


Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                885                 890                 895


Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            900                 905                 910


Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
```

```
        915                 920                 925

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    930                 935                 940

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
945                 950                 955                 960

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                965                 970                 975

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
            980                 985                 990

Asn Gly Leu Phe Gly Asn Leu Ile  Ala Leu Ser Leu Gly  Leu Thr Pro
        995                 1000                1005

Asn Phe  Lys Ser Asn Phe Asp  Leu Ala Glu Asp Ala  Lys Leu Gln
    1010                1015                1020

Leu Ser  Lys Asp Thr Tyr Asp  Asp Asp Leu Asp Asn  Leu Leu Ala
    1025                1030                1035

Gln Ile  Gly Asp Gln Tyr Ala  Asp Leu Phe Leu Ala  Ala Lys Asn
    1040                1045                1050

Leu Ser  Asp Ala Ile Leu Leu  Ser Asp Ile Leu Arg  Val Asn Thr
    1055                1060                1065

Glu Ile  Thr Lys Ala Pro Leu  Ser Ala Ser Met Ile  Lys Arg Tyr
    1070                1075                1080

Asp Glu  His His Gln Asp Leu  Thr Leu Leu Lys Ala  Leu Val Arg
    1085                1090                1095

Gln Gln  Leu Pro Glu Lys Tyr  Lys Glu Ile Phe Phe  Asp Gln Ser
    1100                1105                1110

Lys Asn  Gly Tyr Ala Gly Tyr  Ile Asp Gly Gly Ala  Ser Gln Glu
    1115                1120                1125
```

```
Glu Phe  Tyr Lys Phe Ile Lys  Pro Ile Leu Glu Lys  Met Asp Gly
    1130                1135                1140

Thr Glu  Glu Leu Leu Val Lys  Leu Asn Arg Glu Asp  Leu Leu Arg
    1145                1150                1155

Lys Gln  Arg Thr Phe Asp Asn  Gly Ser Ile Pro His  Gln Ile His
    1160                1165                1170

Leu Gly  Glu Leu His Ala Ile  Leu Arg Arg Gln Glu  Asp Phe Tyr
    1175                1180                1185

Pro Phe  Leu Lys Asp Asn Arg  Glu Lys Ile Glu Lys  Ile Leu Thr
    1190                1195                1200

Phe Arg  Ile Pro Tyr Tyr Val  Gly Pro Leu Ala Arg  Gly Asn Ser
    1205                1210                1215

Arg Phe  Ala Trp Met Thr Arg  Lys Ser Glu Glu Thr  Ile Thr Pro
    1220                1225                1230

Trp Asn  Phe Glu Glu Val Val  Asp Lys Gly Ala Ser  Ala Gln Ser
    1235                1240                1245

Phe Ile  Glu Arg Met Thr Asn  Phe Asp Lys Asn Leu  Pro Asn Glu
    1250                1255                1260

Lys Val  Leu Pro Lys His Ser  Leu Leu Tyr Glu Tyr  Phe Thr Val
    1265                1270                1275

Tyr Asn  Glu Leu Thr Lys Val  Lys Tyr Val Thr Glu  Gly Met Arg
    1280                1285                1290

Lys Pro  Ala Phe Leu Ser Gly  Glu Gln Lys Lys Ala  Ile Val Asp
    1295                1300                1305

Leu Leu  Phe Lys Thr Asn Arg  Lys Val Thr Val Lys  Gln Leu Lys
    1310                1315                1320

Glu Asp  Tyr Phe Lys Lys Ile  Glu Cys Phe Asp Ser  Val Glu Ile
```

```
    1325                1330                1335

Ser Gly  Val Glu Asp Arg Phe  Asn Ala Ser Leu Gly  Thr Tyr His
    1340                1345                1350

Asp Leu  Leu Lys Ile Ile Lys  Asp Lys Asp Phe Leu  Asp Asn Glu
    1355                1360                1365

Glu Asn  Glu Asp Ile Leu Glu  Asp Ile Val Leu Thr  Leu Thr Leu
    1370                1375                1380

Phe Glu  Asp Arg Glu Met Ile  Glu Glu Arg Leu Lys  Thr Tyr Ala
    1385                1390                1395

His Leu  Phe Asp Asp Lys Val  Met Lys Gln Leu Lys  Arg Arg Arg
    1400                1405                1410

Tyr Thr  Gly Trp Gly Arg Leu  Ser Arg Lys Leu Ile  Asn Gly Ile
    1415                1420                1425

Arg Asp  Lys Gln Ser Gly Lys  Thr Ile Leu Asp Phe  Leu Lys Ser
    1430                1435                1440

Asp Gly  Phe Ala Asn Arg Asn  Phe Met Gln Leu Ile  His Asp Asp
    1445                1450                1455

Ser Leu  Thr Phe Lys Glu Asp  Ile Gln Lys Ala Gln  Val Ser Gly
    1460                1465                1470

Gln Gly  Asp Ser Leu His Glu  His Ile Ala Asn Leu  Ala Gly Ser
    1475                1480                1485

Pro Ala  Ile Lys Lys Gly Ile  Leu Gln Thr Val Lys  Val Val Asp
    1490                1495                1500

Glu Leu  Val Lys Val Met Gly  Arg His Lys Pro Glu  Asn Ile Val
    1505                1510                1515

Ile Glu  Met Ala Arg Glu Asn  Gln Thr Thr Gln Lys  Gly Gln Lys
    1520                1525                1530
```

```
Asn Ser  Arg Glu Arg Met Lys  Arg Ile Glu Glu Gly  Ile Lys Glu
    1535                1540                1545

Leu Gly  Ser Gln Ile Leu Lys  Glu His Pro Val Glu  Asn Thr Gln
    1550                1555                1560

Leu Gln  Asn Glu Lys Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg
    1565                1570                1575

Asp Met  Tyr Val Asp Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp
    1580                1585                1590

Tyr Asp  Val Asp His Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp
    1595                1600                1605

Ser Ile  Asp Asn Lys Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly
    1610                1615                1620

Lys Ser  Asp Asn Val Pro Ser  Glu Glu Val Val Lys  Lys Met Lys
    1625                1630                1635

Asn Tyr  Trp Arg Gln Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg
    1640                1645                1650

Lys Phe  Asp Asn Leu Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu
    1655                1660                1665

Leu Asp  Lys Ala Gly Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg
    1670                1675                1680

Gln Ile  Thr Lys His Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn
    1685                1690                1695

Thr Lys  Tyr Asp Glu Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val
    1700                1705                1710

Ile Thr  Leu Lys Ser Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe
    1715                1720                1725

Gln Phe  Tyr Lys Val Arg Glu  Ile Asn Asn Tyr His  His Ala His
```

```
    1730                1735                1740

Asp Ala  Tyr Leu Asn Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys
    1745                1750                1755

Tyr Pro  Lys Leu Glu Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val
    1760                1765                1770

Tyr Asp  Val Arg Lys Met Ile  Ala Lys Ser Glu Gln  Glu Gly Ala
    1775                1780                1785

Asp Lys  Arg Thr Ala Asp Gly  Ser Glu Phe Glu Ser  Pro Lys Lys
    1790                1795                1800

Lys Arg  Lys Val
    1805
```

<210> 333
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: ABE8.8-m polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(1566)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1567)..(1588)
<223> Bipartite nuclear localization sequence

<400> 333

```
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 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 Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220
```

```
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
            420                 425                 430
```

```
Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655
```

```
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860
```

```
Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880


Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895


Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910


Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925


Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940


Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960


Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975


Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990


Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005


Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020


Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035


Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050


Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065


Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080
```

```
Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275
```

```
Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485
```

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1520 1525 1530

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 334
<211> 1786
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: ABE8.8-d polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(365)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (366)..(397)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (398)..(1764)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1765)..(1786)
<223> Bipartite nuclear localization sequence

<400> 334

```
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 Trp Asp Glu Arg Glu Val Pro Val Gly Ala
            20                  25                  30


Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
        35                  40                  45


Ile Gly Arg 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 Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95


Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp 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 Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
```

```
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
    290                 295                 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365
```

```
Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
385                 390                 395                 400

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                405                 410                 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            420                 425                 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        435                 440                 445

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

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465                 470                 475                 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                485                 490                 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            500                 505                 510

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        515                 520                 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    530                 535                 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545                 550                 555                 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                565                 570                 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
```

```
            580                 585                 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        595                 600                 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    610                 615                 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625                 630                 635                 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                645                 650                 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            660                 665                 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        675                 680                 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    690                 695                 700

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
705                 710                 715                 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                725                 730                 735

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

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        755                 760                 765

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    770                 775                 780

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
785                 790                 795                 800
```

```
Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                805                 810                 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            820                 825                 830

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        835                 840                 845

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    850                 855                 860

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
865                 870                 875                 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                885                 890                 895

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            900                 905                 910

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        915                 920                 925

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    930                 935                 940

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945                 950                 955                 960

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                965                 970                 975

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            980                 985                 990

Leu Leu Lys Ile Ile Lys Asp Lys  Asp Phe Leu Asp Asn  Glu Glu Asn
        995                 1000                1005

Glu Asp  Ile Leu Glu Asp Ile  Val Leu Thr Leu Thr  Leu Phe Glu
```

```
    1010                1015                1020

Asp Arg  Glu Met Ile Glu Glu  Arg Leu Lys Thr Tyr  Ala His Leu
    1025                1030                1035

Phe Asp  Asp Lys Val Met Lys  Gln Leu Lys Arg Arg  Arg Tyr Thr
    1040                1045                1050

Gly Trp  Gly Arg Leu Ser Arg  Lys Leu Ile Asn Gly  Ile Arg Asp
    1055                1060                1065

Lys Gln  Ser Gly Lys Thr Ile  Leu Asp Phe Leu Lys  Ser Asp Gly
    1070                1075                1080

Phe Ala  Asn Arg Asn Phe Met  Gln Leu Ile His Asp  Asp Ser Leu
    1085                1090                1095

Thr Phe  Lys Glu Asp Ile Gln  Lys Ala Gln Val Ser  Gly Gln Gly
    1100                1105                1110

Asp Ser  Leu His Glu His Ile  Ala Asn Leu Ala Gly  Ser Pro Ala
    1115                1120                1125

Ile Lys  Lys Gly Ile Leu Gln  Thr Val Lys Val Val  Asp Glu Leu
    1130                1135                1140

Val Lys  Val Met Gly Arg His  Lys Pro Glu Asn Ile  Val Ile Glu
    1145                1150                1155

Met Ala  Arg Glu Asn Gln Thr  Thr Gln Lys Gly Gln  Lys Asn Ser
    1160                1165                1170

Arg Glu  Arg Met Lys Arg Ile  Glu Glu Gly Ile Lys  Glu Leu Gly
    1175                1180                1185

Ser Gln  Ile Leu Lys Glu His  Pro Val Glu Asn Thr  Gln Leu Gln
    1190                1195                1200

Asn Glu  Lys Leu Tyr Leu Tyr  Tyr Leu Gln Asn Gly  Arg Asp Met
    1205                1210                1215
```

```
Tyr Val  Asp Gln Glu Leu Asp  Ile Asn Arg Leu Ser  Asp Tyr Asp
    1220                1225                1230

Val Asp  His Ile Val Pro Gln  Ser Phe Leu Lys Asp  Asp Ser Ile
    1235                1240                1245

Asp Asn  Lys Val Leu Thr Arg  Ser Asp Lys Asn Arg  Gly Lys Ser
    1250                1255                1260

Asp Asn  Val Pro Ser Glu Glu  Val Val Lys Lys Met  Lys Asn Tyr
    1265                1270                1275

Trp Arg  Gln Leu Leu Asn Ala  Lys Leu Ile Thr Gln  Arg Lys Phe
    1280                1285                1290

Asp Asn  Leu Thr Lys Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp
    1295                1300                1305

Lys Ala  Gly Phe Ile Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile
    1310                1315                1320

Thr Lys  His Val Ala Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys
    1325                1330                1335

Tyr Asp  Glu Asn Asp Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr
    1340                1345                1350

Leu Lys  Ser Lys Leu Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe
    1355                1360                1365

Tyr Lys  Val Arg Glu Ile Asn  Asn Tyr His His Ala  His Asp Ala
    1370                1375                1380

Tyr Leu  Asn Ala Val Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro
    1385                1390                1395

Lys Leu  Glu Ser Glu Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp
    1400                1405                1410

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
```

```
    1415                1420                1425

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1430                1435                1440

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1445                1450                1455

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1460                1465                1470

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1475                1480                1485

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1490                1495                1500

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1505                1510                1515

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1520                1525                1530

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1535                1540                1545

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1550                1555                1560

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1565                1570                1575

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1580                1585                1590

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1595                1600                1605

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1610                1615                1620
```

```
Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1625                1630                1635


Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1640                1645                1650


Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1655                1660                1665


Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1670                1675                1680


Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1685                1690                1695


Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1700                1705                1710


Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1715                1720                1725


Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1730                1735                1740


Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1745                1750                1755


Ser Gln  Leu Gly Gly Asp Glu  Gly Ala Asp Lys Arg  Thr Ala Asp
    1760                1765                1770


Gly Ser  Glu Phe Glu Ser Pro  Lys Lys Lys Arg Lys  Val
    1775                1780                1785


<210>  335
<211>  1588
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: ABE8.13-m polypeptide
       sequence
```

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(1566)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1567)..(1588)
<223> Bipartite nuclear localization sequence

<400> 335

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 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 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 Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
```

```
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560
```

```
Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
                565                 570                 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
```

```
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990
```

```
Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005


Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020


Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035


Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050


Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065


Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080


Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095


Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110


Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125


Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140


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


Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170


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


Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
```

```
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395
```

```
Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg
    1520                1525                1530

Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 336

<211> 1786
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: ABE8.13-d polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(365)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (366)..(397)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (398)..(1764)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1765)..(1786)
<223> Bipartite nuclear localization sequence

<400> 336

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 Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg 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 Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp 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 Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270
```

```
Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
    290                 295                 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
385                 390                 395                 400

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                405                 410                 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            420                 425                 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        435                 440                 445

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

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465                 470                 475                 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
```

```
                485                 490                 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            500                 505                 510

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        515                 520                 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    530                 535                 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545                 550                 555                 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                565                 570                 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            580                 585                 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        595                 600                 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    610                 615                 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625                 630                 635                 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                645                 650                 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            660                 665                 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        675                 680                 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    690                 695                 700
```

```
Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
705                 710                 715                 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                725                 730                 735

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

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        755                 760                 765

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    770                 775                 780

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
785                 790                 795                 800

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                805                 810                 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            820                 825                 830

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        835                 840                 845

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    850                 855                 860

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
865                 870                 875                 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                885                 890                 895

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            900                 905                 910

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
```

```
        915                 920                 925

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    930                 935                 940

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945                 950                 955                 960

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                965                 970                 975

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            980                 985                 990

Leu Leu Lys Ile Ile Lys Asp Lys  Asp Phe Leu Asp Asn  Glu Glu Asn
        995                 1000                1005

Glu Asp  Ile Leu Glu Asp Ile  Val Leu Thr Leu Thr  Leu Phe Glu
    1010                1015                1020

Asp Arg  Glu Met Ile Glu Glu  Arg Leu Lys Thr Tyr  Ala His Leu
    1025                1030                1035

Phe Asp  Asp Lys Val Met Lys  Gln Leu Lys Arg Arg  Arg Tyr Thr
    1040                1045                1050

Gly Trp  Gly Arg Leu Ser Arg  Lys Leu Ile Asn Gly  Ile Arg Asp
    1055                1060                1065

Lys Gln  Ser Gly Lys Thr Ile  Leu Asp Phe Leu Lys  Ser Asp Gly
    1070                1075                1080

Phe Ala  Asn Arg Asn Phe Met  Gln Leu Ile His Asp  Asp Ser Leu
    1085                1090                1095

Thr Phe  Lys Glu Asp Ile Gln  Lys Ala Gln Val Ser  Gly Gln Gly
    1100                1105                1110

Asp Ser  Leu His Glu His Ile  Ala Asn Leu Ala Gly  Ser Pro Ala
    1115                1120                1125
```

```
Ile Lys  Lys Gly Ile Leu Gln  Thr Val Lys Val Val  Asp Glu Leu
    1130                1135                1140

Val Lys  Val Met Gly Arg His  Lys Pro Glu Asn Ile  Val Ile Glu
    1145                1150                1155

Met Ala  Arg Glu Asn Gln Thr  Thr Gln Lys Gly Gln  Lys Asn Ser
    1160                1165                1170

Arg Glu  Arg Met Lys Arg Ile  Glu Glu Gly Ile Lys  Glu Leu Gly
    1175                1180                1185

Ser Gln  Ile Leu Lys Glu His  Pro Val Glu Asn Thr  Gln Leu Gln
    1190                1195                1200

Asn Glu  Lys Leu Tyr Leu Tyr  Tyr Leu Gln Asn Gly  Arg Asp Met
    1205                1210                1215

Tyr Val  Asp Gln Glu Leu Asp  Ile Asn Arg Leu Ser  Asp Tyr Asp
    1220                1225                1230

Val Asp  His Ile Val Pro Gln  Ser Phe Leu Lys Asp  Asp Ser Ile
    1235                1240                1245

Asp Asn  Lys Val Leu Thr Arg  Ser Asp Lys Asn Arg  Gly Lys Ser
    1250                1255                1260

Asp Asn  Val Pro Ser Glu Glu  Val Val Lys Lys Met  Lys Asn Tyr
    1265                1270                1275

Trp Arg  Gln Leu Leu Asn Ala  Lys Leu Ile Thr Gln  Arg Lys Phe
    1280                1285                1290

Asp Asn  Leu Thr Lys Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp
    1295                1300                1305

Lys Ala  Gly Phe Ile Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile
    1310                1315                1320

Thr Lys  His Val Ala Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys
```

```
    1325                1330                1335

Tyr Asp  Glu Asn Asp Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr
    1340                1345                1350

Leu Lys  Ser Lys Leu Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe
    1355                1360                1365

Tyr Lys  Val Arg Glu Ile Asn  Asn Tyr His His Ala  His Asp Ala
    1370                1375                1380

Tyr Leu  Asn Ala Val Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro
    1385                1390                1395

Lys Leu  Glu Ser Glu Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp
    1400                1405                1410

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1415                1420                1425

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1430                1435                1440

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1445                1450                1455

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1460                1465                1470

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1475                1480                1485

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1490                1495                1500

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1505                1510                1515

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1520                1525                1530
```

```
Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1535                1540                1545

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1550                1555                1560

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1565                1570                1575

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1580                1585                1590

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1595                1600                1605

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1610                1615                1620

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1625                1630                1635

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1640                1645                1650

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1655                1660                1665

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1670                1675                1680

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1685                1690                1695

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1700                1705                1710

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1715                1720                1725

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
```

1730 1735 1740

Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1745 1750 1755

Ser Gln Leu Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp
1760 1765 1770

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1775 1780 1785

<210> 337
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: ABE8.17-m polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(1566)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1567)..(1588)
<223> Bipartite nuclear localization sequence

<400> 337

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 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 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 Tyr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
```

```
               245                 250                 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460
```

```
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480


Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495


Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510


Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525


Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540


Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560


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


Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590


Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605


Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620


Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640


Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655


Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670


Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
```

```
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895
```

```
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
```

```
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305
```

```
Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
```

```
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Arg
    1520                1525                1530

Tyr Thr  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 338
<211> 1786
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: ABE8.17-d polypeptide sequence

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(365)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (366)..(397)
<223> Linker sequence

<220>

<221> MISC_FEATURE
<222> (398)..(1764)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1765)..(1786)
<223> Bipartite nuclear localization sequence

<400> 338

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 Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg 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 Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp 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 Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

```
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270

Leu Ile Asp Ala Thr Leu Tyr Ser Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
    290                 295                 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Arg
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
```

```
385                 390                 395                 400

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                405                 410                 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            420                 425                 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        435                 440                 445

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

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465                 470                 475                 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                485                 490                 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            500                 505                 510

Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        515                 520                 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    530                 535                 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545                 550                 555                 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                565                 570                 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            580                 585                 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        595                 600                 605
```

```
Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    610                 615                 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625                 630                 635                 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                645                 650                 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            660                 665                 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        675                 680                 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    690                 695                 700

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
705                 710                 715                 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
                725                 730                 735

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

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        755                 760                 765

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    770                 775                 780

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
785                 790                 795                 800

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                805                 810                 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
```

```
            820                 825                 830

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        835                 840                 845

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    850                 855                 860

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
865                 870                 875                 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                885                 890                 895

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            900                 905                 910

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        915                 920                 925

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    930                 935                 940

Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945                 950                 955                 960

Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                965                 970                 975

Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            980                 985                 990

Leu Leu Lys Ile Ile Lys Asp Lys  Asp Phe Leu Asp Asn  Glu Glu Asn
        995                 1000                1005

Glu Asp  Ile Leu Glu Asp Ile  Val Leu Thr Leu Thr  Leu Phe Glu
    1010                1015                1020

Asp Arg  Glu Met Ile Glu Glu  Arg Leu Lys Thr Tyr  Ala His Leu
    1025                1030                1035
```

```
Phe Asp  Asp Lys Val Met Lys  Gln Leu Lys Arg Arg  Arg Tyr Thr
    1040                1045                1050

Gly Trp  Gly Arg Leu Ser Arg  Lys Leu Ile Asn Gly  Ile Arg Asp
    1055                1060                1065

Lys Gln  Ser Gly Lys Thr Ile  Leu Asp Phe Leu Lys  Ser Asp Gly
    1070                1075                1080

Phe Ala  Asn Arg Asn Phe Met  Gln Leu Ile His Asp  Asp Ser Leu
    1085                1090                1095

Thr Phe  Lys Glu Asp Ile Gln  Lys Ala Gln Val Ser  Gly Gln Gly
    1100                1105                1110

Asp Ser  Leu His Glu His Ile  Ala Asn Leu Ala Gly  Ser Pro Ala
    1115                1120                1125

Ile Lys  Lys Gly Ile Leu Gln  Thr Val Lys Val Val  Asp Glu Leu
    1130                1135                1140

Val Lys  Val Met Gly Arg His  Lys Pro Glu Asn Ile  Val Ile Glu
    1145                1150                1155

Met Ala  Arg Glu Asn Gln Thr  Thr Gln Lys Gly Gln  Lys Asn Ser
    1160                1165                1170

Arg Glu  Arg Met Lys Arg Ile  Glu Glu Gly Ile Lys  Glu Leu Gly
    1175                1180                1185

Ser Gln  Ile Leu Lys Glu His  Pro Val Glu Asn Thr  Gln Leu Gln
    1190                1195                1200

Asn Glu  Lys Leu Tyr Leu Tyr  Tyr Leu Gln Asn Gly  Arg Asp Met
    1205                1210                1215

Tyr Val  Asp Gln Glu Leu Asp  Ile Asn Arg Leu Ser  Asp Tyr Asp
    1220                1225                1230

Val Asp  His Ile Val Pro Gln  Ser Phe Leu Lys Asp  Asp Ser Ile
```

```
    1235                1240                1245

Asp Asn  Lys Val Leu Thr Arg  Ser Asp Lys Asn Arg  Gly Lys Ser
    1250                1255                1260

Asp Asn  Val Pro Ser Glu Glu  Val Val Lys Lys Met  Lys Asn Tyr
    1265                1270                1275

Trp Arg  Gln Leu Leu Asn Ala  Lys Leu Ile Thr Gln  Arg Lys Phe
    1280                1285                1290

Asp Asn  Leu Thr Lys Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp
    1295                1300                1305

Lys Ala  Gly Phe Ile Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile
    1310                1315                1320

Thr Lys  His Val Ala Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys
    1325                1330                1335

Tyr Asp  Glu Asn Asp Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr
    1340                1345                1350

Leu Lys  Ser Lys Leu Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe
    1355                1360                1365

Tyr Lys  Val Arg Glu Ile Asn  Asn Tyr His His Ala  His Asp Ala
    1370                1375                1380

Tyr Leu  Asn Ala Val Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro
    1385                1390                1395

Lys Leu  Glu Ser Glu Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp
    1400                1405                1410

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1415                1420                1425

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1430                1435                1440
```

```
Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1445                1450                1455

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1460                1465                1470

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1475                1480                1485

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1490                1495                1500

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1505                1510                1515

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1520                1525                1530

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1535                1540                1545

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
    1550                1555                1560

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1565                1570                1575

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1580                1585                1590

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1595                1600                1605

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1610                1615                1620

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1625                1630                1635

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
```

```
    1640                1645                1650

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1655                1660                1665

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1670                1675                1680

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1685                1690                1695

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1700                1705                1710

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1715                1720                1725

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1730                1735                1740

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1745                1750                1755

Ser Gln  Leu Gly Gly Asp Glu  Gly Ala Asp Lys Arg  Thr Ala Asp
    1760                1765                1770

Gly Ser  Glu Phe Glu Ser Pro  Lys Lys Lys Arg Lys  Val
    1775                1780                1785


<210>  339
<211>  1588
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: ABE8.20-m polypeptide
       sequence


<220>
<221>  MISC_FEATURE
<222>  (1)..(167)
<223>  Adenosine deaminse sequence
```

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(1566)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1567)..(1588)
<223> Bipartite nuclear localization sequence

<400> 339

```
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 Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365
```

```
Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
```

```
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800
```

```
Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
```

```
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215
```

```
Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Asp Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Gly Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
```

1415 1420 1425

Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
1430 1435 1440

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg
1520 1525 1530

Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 340
<211> 1786
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: ABE8.20-d polypeptide

sequence

<220>
<221> MISC_FEATURE
<222> (1)..(167)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (168)..(199)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (200)..(365)
<223> Adenosine deaminase sequence

<220>
<221> MISC_FEATURE
<222> (366)..(397)
<223> Linker sequence

<220>
<221> MISC_FEATURE
<222> (398)..(1764)
<223> Sequence derived from Cas9

<220>
<221> MISC_FEATURE
<222> (1765)..(1786)
<223> Bipartite nuclear localization sequence

<400> 340

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 Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
35 40 45

Ile Gly Arg 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 Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp 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 Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270

Leu Tyr Asp Ala Thr Leu Tyr Ser Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
```

```
    290                 295                 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg Arg
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys
385                 390                 395                 400

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
                405                 410                 415

Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
            420                 425                 430

Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
        435                 440                 445

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

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
465                 470                 475                 480

Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His Arg
                485                 490                 495

Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg His
            500                 505                 510
```

```
Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys Tyr
        515                 520                 525

Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp Lys
    530                 535                 540

Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Phe
545                 550                 555                 560

Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp
                565                 570                 575

Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe
            580                 585                 590

Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile Leu
        595                 600                 605

Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln
    610                 615                 620

Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu
625                 630                 635                 640

Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu
                645                 650                 655

Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp
            660                 665                 670

Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala
        675                 680                 685

Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu Arg Val
    690                 695                 700

Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile Lys Arg
705                 710                 715                 720

Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu Val Arg
```

```
                725                 730                 735

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

Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe
        755                 760                 765

Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr Glu Glu
    770                 775                 780

Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr
785                 790                 795                 800

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu Leu His
                805                 810                 815

Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn
            820                 825                 830

Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
        835                 840                 845

Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr Arg Lys
    850                 855                 860

Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val Asp Lys
865                 870                 875                 880

Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe Asp Lys
                885                 890                 895

Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu Tyr Glu
            900                 905                 910

Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Thr Glu
        915                 920                 925

Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile
    930                 935                 940
```

```
Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys Gln Leu
945                 950                 955                 960


Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val Glu Ile
                965                 970                 975


Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr His Asp
            980                 985                 990


Leu Leu Lys Ile Ile Lys Asp Lys  Asp Phe Leu Asp Asn  Glu Glu Asn
        995                 1000                1005


Glu Asp  Ile Leu Glu Asp Ile  Val Leu Thr Leu Thr  Leu Phe Glu
    1010                1015                1020


Asp Arg  Glu Met Ile Glu Glu  Arg Leu Lys Thr Tyr  Ala His Leu
    1025                1030                1035


Phe Asp  Asp Lys Val Met Lys  Gln Leu Lys Arg Arg  Arg Tyr Thr
    1040                1045                1050


Gly Trp  Gly Arg Leu Ser Arg  Lys Leu Ile Asn Gly  Ile Arg Asp
    1055                1060                1065


Lys Gln  Ser Gly Lys Thr Ile  Leu Asp Phe Leu Lys  Ser Asp Gly
    1070                1075                1080


Phe Ala  Asn Arg Asn Phe Met  Gln Leu Ile His Asp  Asp Ser Leu
    1085                1090                1095


Thr Phe  Lys Glu Asp Ile Gln  Lys Ala Gln Val Ser  Gly Gln Gly
    1100                1105                1110


Asp Ser  Leu His Glu His Ile  Ala Asn Leu Ala Gly  Ser Pro Ala
    1115                1120                1125


Ile Lys  Lys Gly Ile Leu Gln  Thr Val Lys Val Val  Asp Glu Leu
    1130                1135                1140


Val Lys  Val Met Gly Arg His  Lys Pro Glu Asn Ile  Val Ile Glu
```

```
    1145                1150                1155

Met Ala  Arg Glu Asn Gln Thr  Thr Gln Lys Gly Gln  Lys Asn Ser
    1160                1165                1170

Arg Glu  Arg Met Lys Arg Ile  Glu Glu Gly Ile Lys  Glu Leu Gly
    1175                1180                1185

Ser Gln  Ile Leu Lys Glu His  Pro Val Glu Asn Thr  Gln Leu Gln
    1190                1195                1200

Asn Glu  Lys Leu Tyr Leu Tyr  Tyr Leu Gln Asn Gly  Arg Asp Met
    1205                1210                1215

Tyr Val  Asp Gln Glu Leu Asp  Ile Asn Arg Leu Ser  Asp Tyr Asp
    1220                1225                1230

Val Asp  His Ile Val Pro Gln  Ser Phe Leu Lys Asp  Asp Ser Ile
    1235                1240                1245

Asp Asn  Lys Val Leu Thr Arg  Ser Asp Lys Asn Arg  Gly Lys Ser
    1250                1255                1260

Asp Asn  Val Pro Ser Glu Glu  Val Val Lys Lys Met  Lys Asn Tyr
    1265                1270                1275

Trp Arg  Gln Leu Leu Asn Ala  Lys Leu Ile Thr Gln  Arg Lys Phe
    1280                1285                1290

Asp Asn  Leu Thr Lys Ala Glu  Arg Gly Gly Leu Ser  Glu Leu Asp
    1295                1300                1305

Lys Ala  Gly Phe Ile Lys Arg  Gln Leu Val Glu Thr  Arg Gln Ile
    1310                1315                1320

Thr Lys  His Val Ala Gln Ile  Leu Asp Ser Arg Met  Asn Thr Lys
    1325                1330                1335

Tyr Asp  Glu Asn Asp Lys Leu  Ile Arg Glu Val Lys  Val Ile Thr
    1340                1345                1350
```

```
Leu Lys  Ser Lys Leu Val Ser  Asp Phe Arg Lys Asp  Phe Gln Phe
    1355                1360                1365

Tyr Lys  Val Arg Glu Ile Asn  Asn Tyr His His Ala  His Asp Ala
    1370                1375                1380

Tyr Leu  Asn Ala Val Val Gly  Thr Ala Leu Ile Lys  Lys Tyr Pro
    1385                1390                1395

Lys Leu  Glu Ser Glu Phe Val  Tyr Gly Asp Tyr Lys  Val Tyr Asp
    1400                1405                1410

Val Arg  Lys Met Ile Ala Lys  Ser Glu Gln Glu Ile  Gly Lys Ala
    1415                1420                1425

Thr Ala  Lys Tyr Phe Phe Tyr  Ser Asn Ile Met Asn  Phe Phe Lys
    1430                1435                1440

Thr Glu  Ile Thr Leu Ala Asn  Gly Glu Ile Arg Lys  Arg Pro Leu
    1445                1450                1455

Ile Glu  Thr Asn Gly Glu Thr  Gly Glu Ile Val Trp  Asp Lys Gly
    1460                1465                1470

Arg Asp  Phe Ala Thr Val Arg  Lys Val Leu Ser Met  Pro Gln Val
    1475                1480                1485

Asn Ile  Val Lys Lys Thr Glu  Val Gln Thr Gly Gly  Phe Ser Lys
    1490                1495                1500

Glu Ser  Ile Leu Pro Lys Arg  Asn Ser Asp Lys Leu  Ile Ala Arg
    1505                1510                1515

Lys Lys  Asp Trp Asp Pro Lys  Lys Tyr Gly Gly Phe  Asp Ser Pro
    1520                1525                1530

Thr Val  Ala Tyr Ser Val Leu  Val Val Ala Lys Val  Glu Lys Gly
    1535                1540                1545

Lys Ser  Lys Lys Leu Lys Ser  Val Lys Glu Leu Leu  Gly Ile Thr
```

```
    1550                1555                1560

Ile Met  Glu Arg Ser Ser Phe  Glu Lys Asn Pro Ile  Asp Phe Leu
    1565                1570                1575

Glu Ala  Lys Gly Tyr Lys Glu  Val Lys Lys Asp Leu  Ile Ile Lys
    1580                1585                1590

Leu Pro  Lys Tyr Ser Leu Phe  Glu Leu Glu Asn Gly  Arg Lys Arg
    1595                1600                1605

Met Leu  Ala Ser Ala Gly Glu  Leu Gln Lys Gly Asn  Glu Leu Ala
    1610                1615                1620

Leu Pro  Ser Lys Tyr Val Asn  Phe Leu Tyr Leu Ala  Ser His Tyr
    1625                1630                1635

Glu Lys  Leu Lys Gly Ser Pro  Glu Asp Asn Glu Gln  Lys Gln Leu
    1640                1645                1650

Phe Val  Glu Gln His Lys His  Tyr Leu Asp Glu Ile  Ile Glu Gln
    1655                1660                1665

Ile Ser  Glu Phe Ser Lys Arg  Val Ile Leu Ala Asp  Ala Asn Leu
    1670                1675                1680

Asp Lys  Val Leu Ser Ala Tyr  Asn Lys His Arg Asp  Lys Pro Ile
    1685                1690                1695

Arg Glu  Gln Ala Glu Asn Ile  Ile His Leu Phe Thr  Leu Thr Asn
    1700                1705                1710

Leu Gly  Ala Pro Ala Ala Phe  Lys Tyr Phe Asp Thr  Thr Ile Asp
    1715                1720                1725

Arg Lys  Arg Tyr Thr Ser Thr  Lys Glu Val Leu Asp  Ala Thr Leu
    1730                1735                1740

Ile His  Gln Ser Ile Thr Gly  Leu Tyr Glu Thr Arg  Ile Asp Leu
    1745                1750                1755
```

Ser Gln Leu Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp
    1760   1765   1770

Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
    1775   1780   1785

<210> 341
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 01. monoABE8.1_bpNLS + Y147T
       polypeptide sequence

<400> 341

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 Ala Lys Thr Gly
            100   105   110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Thr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350
```

```
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
                565                 570                 575
```

```
Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780
```

```
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800


Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815


Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830


Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845


Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860


Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880


Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895


Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910


Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925


Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940


Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960


Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975


Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990


Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005
```

```
Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200
```

```
Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                 1210                 1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                 1225                 1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                 1240                 1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                 1255                 1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                 1270                 1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                 1285                 1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                 1300                 1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                 1315                 1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                 1330                 1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                 1345                 1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                 1360                 1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                 1375                 1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                 1390                 1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                 1405                 1410
```

```
Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 342
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 02. monoABE8.1_bpNLS + Y147R polypeptide sequence

<400> 342

```
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 Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Arg Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190
```

```
Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
```

```
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620
```

```
Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
```

```
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050
```

```
Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
```

```
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455
```

```
Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470


Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485


Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500


Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515


Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530


Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545


Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560


Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575


Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 343
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 03. monoABE8.1_bpNLS + Q154S polypeptide sequence

<400> 343

```
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 Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Tyr Phe Phe Arg Met Pro Arg Ser Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240
```

```
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
                245                 250                 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460
```

```
Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670
```

```
Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895
```

```
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095
```

```
Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305
```

```
Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500
```

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
    1505   1510   1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
    1520   1525   1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
    1535   1540   1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
    1550   1555   1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
    1565   1570   1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
    1580   1585

<210> 344
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 04. monoABE8.1_bpNLS + Y123H
       polypeptide sequence

<400> 344

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 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 Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
```

```
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510
```

```
Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
```

```
               725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940
```

```
Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960


Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975


Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990


Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005


Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020


Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035


Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050


Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065


Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080


Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095


Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110


Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125


Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140


Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val Ile Thr  Leu Lys Ser
```

```
    1145                1150                1155

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350
```

```
Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
```

```
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 345
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 05. monoABE8.1_bpNLS + V82S polypeptide sequence

<400> 345

```
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 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 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 Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350
```

```
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560
```

```
Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
                565                 570                 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780
```

```
Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990
```

```
Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005


Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020


Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035


Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050


Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065


Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080


Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095


Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110


Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125


Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140


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


Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170


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


Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200
```

```
Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                 1210                 1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                 1225                 1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                 1240                 1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                 1255                 1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                 1270                 1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                 1285                 1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                 1300                 1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                 1315                 1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                 1330                 1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                 1345                 1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                 1360                 1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                 1375                 1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                 1390                 1395
```

```
Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 346
<211> 1588

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 06. monoABE8.1_bpNLS + T166R polypeptide sequence

<400> 346

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 Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Arg Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

```
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400
```

```
Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
```

```
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830
```

```
Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
```

```
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245
```

```
Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
```

```
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 347
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 07. monoABE8.1_bpNLS + Q154R polypeptide sequence

<400> 347

```
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 Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Tyr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240
```

```
Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
                245                 250                 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445
```

```
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670
```

```
Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880
```

```
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
               885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
           900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
               965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
           980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095
```

```
Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290
```

```
Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                 1300                 1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                 1315                 1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                 1330                 1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                 1345                 1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                 1360                 1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                 1375                 1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                 1390                 1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                 1405                 1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                 1420                 1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                 1435                 1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                 1450                 1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                 1465                 1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                 1480                 1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                 1495                 1500
```

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
    1505   1510   1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
    1520   1525   1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
    1535   1540   1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
    1550   1555   1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
    1565   1570   1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
    1580   1585

<210> 348
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 08. monoABE8.1_bpNLS + Y147R_Q154R_Y123H polypeptide sequence

<400> 348

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 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 Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285
```

```
Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
```

```
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720
```

```
Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
```

```
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140
```

```
Asn Asp  Lys Leu Ile Arg Glu  Val Lys Val Ile Thr  Leu Lys Ser
    1145                1150                1155

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
```

```
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545
```

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
    1550   1555   1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
    1565   1570   1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
    1580   1585

<210> 349
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 09. monoABE8.1_bpNLS + Y147R_Q154R_I76Y polypeptide sequence

<400> 349

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 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 Ala Lys Thr Gly
            100   105   110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 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 Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335
```

```
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560
```

```
Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
                565                 570                 575

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

```
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990
```

```
Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

Arg Glu  Ile Asn Asn Tyr His  His Ala His Asp Ala  Tyr Leu Asn
    1175                1180                1185
```

```
Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395
```

```
Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 350
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 10. monoABE8.1_bpNLS + Y147R_Q154R_T166R polypeptide sequence

<400> 350

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 Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Arg Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
165 170 175

```
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
```

```
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605
```

```
Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
```

```
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035
```

```
Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
```

```
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440
```

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu
1445 1450 1455

Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu
1460 1465 1470

Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1475 1480 1485

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
1490 1495 1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
1505 1510 1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
1520 1525 1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 351
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 11. monoABE8.1_bpNLS + Y147T_Q154R polypeptide sequence

<400> 351

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 Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Thr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220
```

```
Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445
```

```
Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655
```

```
Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880
```

```
Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
               885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
               965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080
```

```
Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290
```

```
Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485
```

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln
    1490   1495   1500

Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala
    1505   1510   1515

Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Gln
    1520   1525   1530

Tyr Arg Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln
    1535   1540   1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
    1550   1555   1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
    1565   1570   1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
    1580   1585

<210> 352
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 12. monoABE8.1_bpNLS +
       Y147T_Q154S polypeptide sequence

<400> 352

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 Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His 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 Thr Phe Phe Arg Met Pro Arg Ser Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
```

```
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495
```

```
Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
```

```
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925
```

```
Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
```

```
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335
```

```
Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
```

1535 1540 1545

Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu
1550 1555 1560

Gly Gly Asp Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu
1565 1570 1575

Phe Glu Ser Pro Lys Lys Lys Arg Lys Val
1580 1585

<210> 353
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 13. monoABE8.1_bpNLS + H123Y123H_Y147R_Q154R_I76Y polypeptide sequence

<400> 353

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 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 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 Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335
```

```
Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540
```

```
Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560


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


Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590


Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
        595                 600                 605


Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620


Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640


Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655


Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670


Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685


Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700


Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720


Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735


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


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

```
Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975
```

```
Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990


Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005


Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020


Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
    1025                1030                1035


Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050


Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065


Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080


Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095


Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110


Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125


Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140


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


Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170


Arg Glu  Ile Asn Asn Tyr His  His Ala His Asp Ala  Tyr Leu Asn
    1175                1180                1185
```

```
Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230

Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380
```

```
Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 354
<211> 1588
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: 14. monoABE8.1_bpNLS + V82S + Q154R polypeptide sequence

<400> 354

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 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 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 Tyr Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys
145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

```
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala
        195                 200                 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
    210                 215                 220

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
225                 230                 235                 240

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

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            260                 265                 270

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        275                 280                 285

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    290                 295                 300

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
305                 310                 315                 320

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                325                 330                 335

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            340                 345                 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        355                 360                 365

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    370                 375                 380
```

```
Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
385                 390                 395                 400

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                405                 410                 415

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

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        435                 440                 445

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    450                 455                 460

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
465                 470                 475                 480

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                485                 490                 495

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            500                 505                 510

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        515                 520                 525

Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    530                 535                 540

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
545                 550                 555                 560

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

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            580                 585                 590

Arg Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile
```

```
        595                 600                 605

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    610                 615                 620

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
625                 630                 635                 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                645                 650                 655

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            660                 665                 670

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        675                 680                 685

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    690                 695                 700

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
705                 710                 715                 720

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                725                 730                 735

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

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

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    770                 775                 780

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
785                 790                 795                 800

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                805                 810                 815
```

```
Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            820                 825                 830

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        835                 840                 845

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    850                 855                 860

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
865                 870                 875                 880

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                885                 890                 895

Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            900                 905                 910

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        915                 920                 925

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    930                 935                 940

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
945                 950                 955                 960

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                965                 970                 975

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            980                 985                 990

Ile Leu Lys Glu His Pro Val Glu  Asn Thr Gln Leu Gln  Asn Glu Lys
        995                 1000                1005

Leu Tyr  Leu Tyr Tyr Leu Gln  Asn Gly Arg Asp Met  Tyr Val Asp
    1010                1015                1020

Gln Glu  Leu Asp Ile Asn Arg  Leu Ser Asp Tyr Asp  Val Asp His
```

```
    1025                1030                1035

Ile Val  Pro Gln Ser Phe Leu  Lys Asp Asp Ser Ile  Asp Asn Lys
    1040                1045                1050

Val Leu  Thr Arg Ser Asp Lys  Asn Arg Gly Lys Ser  Asp Asn Val
    1055                1060                1065

Pro Ser  Glu Glu Val Val Lys  Lys Met Lys Asn Tyr  Trp Arg Gln
    1070                1075                1080

Leu Leu  Asn Ala Lys Leu Ile  Thr Gln Arg Lys Phe  Asp Asn Leu
    1085                1090                1095

Thr Lys  Ala Glu Arg Gly Gly  Leu Ser Glu Leu Asp  Lys Ala Gly
    1100                1105                1110

Phe Ile  Lys Arg Gln Leu Val  Glu Thr Arg Gln Ile  Thr Lys His
    1115                1120                1125

Val Ala  Gln Ile Leu Asp Ser  Arg Met Asn Thr Lys  Tyr Asp Glu
    1130                1135                1140

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

Lys Leu  Val Ser Asp Phe Arg  Lys Asp Phe Gln Phe  Tyr Lys Val
    1160                1165                1170

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

Ala Val  Val Gly Thr Ala Leu  Ile Lys Lys Tyr Pro  Lys Leu Glu
    1190                1195                1200

Ser Glu  Phe Val Tyr Gly Asp  Tyr Lys Val Tyr Asp  Val Arg Lys
    1205                1210                1215

Met Ile  Ala Lys Ser Glu Gln  Glu Ile Gly Lys Ala  Thr Ala Lys
    1220                1225                1230
```

```
Tyr Phe  Phe Tyr Ser Asn Ile  Met Asn Phe Phe Lys  Thr Glu Ile
    1235                1240                1245

Thr Leu  Ala Asn Gly Glu Ile  Arg Lys Arg Pro Leu  Ile Glu Thr
    1250                1255                1260

Asn Gly  Glu Thr Gly Glu Ile  Val Trp Asp Lys Gly  Arg Asp Phe
    1265                1270                1275

Ala Thr  Val Arg Lys Val Leu  Ser Met Pro Gln Val  Asn Ile Val
    1280                1285                1290

Lys Lys  Thr Glu Val Gln Thr  Gly Gly Phe Ser Lys  Glu Ser Ile
    1295                1300                1305

Leu Pro  Lys Arg Asn Ser Asp  Lys Leu Ile Ala Arg  Lys Lys Asp
    1310                1315                1320

Trp Asp  Pro Lys Lys Tyr Gly  Gly Phe Val Ser Pro  Thr Val Ala
    1325                1330                1335

Tyr Ser  Val Leu Val Val Ala  Lys Val Glu Lys Gly  Lys Ser Lys
    1340                1345                1350

Lys Leu  Lys Ser Val Lys Glu  Leu Leu Gly Ile Thr  Ile Met Glu
    1355                1360                1365

Arg Ser  Ser Phe Glu Lys Asn  Pro Ile Asp Phe Leu  Glu Ala Lys
    1370                1375                1380

Gly Tyr  Lys Glu Val Lys Lys  Asp Leu Ile Ile Lys  Leu Pro Lys
    1385                1390                1395

Tyr Ser  Leu Phe Glu Leu Glu  Asn Gly Arg Lys Arg  Met Leu Ala
    1400                1405                1410

Ser Ala  Arg Glu Leu Gln Lys  Gly Asn Glu Leu Ala  Leu Pro Ser
    1415                1420                1425

Lys Tyr  Val Asn Phe Leu Tyr  Leu Ala Ser His Tyr  Glu Lys Leu
```

```
    1430                1435                1440

Lys Gly  Ser Pro Glu Asp Asn  Glu Gln Lys Gln Leu  Phe Val Glu
    1445                1450                1455

Gln His  Lys His Tyr Leu Asp  Glu Ile Ile Glu Gln  Ile Ser Glu
    1460                1465                1470

Phe Ser  Lys Arg Val Ile Leu  Ala Asp Ala Asn Leu  Asp Lys Val
    1475                1480                1485

Leu Ser  Ala Tyr Asn Lys His  Arg Asp Lys Pro Ile  Arg Glu Gln
    1490                1495                1500

Ala Glu  Asn Ile Ile His Leu  Phe Thr Leu Thr Asn  Leu Gly Ala
    1505                1510                1515

Pro Ala  Ala Phe Lys Tyr Phe  Asp Thr Thr Ile Asp  Arg Lys Gln
    1520                1525                1530

Tyr Arg  Ser Thr Lys Glu Val  Leu Asp Ala Thr Leu  Ile His Gln
    1535                1540                1545

Ser Ile  Thr Gly Leu Tyr Glu  Thr Arg Ile Asp Leu  Ser Gln Leu
    1550                1555                1560

Gly Gly  Asp Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1565                1570                1575

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1580                1585
```

<210> 355
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 355

Ser Gly Gly Ser

1

<210> 356
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 356

Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala
1 5 10 15

Thr Pro Glu Ser Ser Gly Gly Ser
20

<210> 357
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 357

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1 5 10 15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
20 25 30

<210> 358
<211> 104
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 358

Gly Gly Ser Gly Gly Ser Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
1 5 10 15

Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
            20   25   30

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
        35   40   45

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
    50   55   60

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
65   70   75   80

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
                85   90   95

Thr Ser Gly Gly Ser Gly Gly Ser
            100

<210> 359
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 359

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1   5   10   15

Ser Glu Ser Ala Thr Pro Glu Ser
            20

<210> 360
<211> 40
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide
       sequence

<400> 360

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1   5   10   15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
            20   25   30

Ser Gly Gly Ser Ser Gly Gly Ser
        35   40

<210> 361
<211> 64
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide
       sequence

<400> 361

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
1   5   10   15

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
            20   25   30

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
        35   40   45

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
    50   55   60

<210> 362
<211> 92
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide
       sequence

<400> 362

Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser
1   5   10   15

Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser
            20   25   30

Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu
        35   40   45

Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser
    50   55   60

Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr
65   70   75   80

Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser
                85   90

<210> 363
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 363

Pro Ala Pro Ala Pro
1   5

<210> 364
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 364

Pro Ala Pro Ala Pro Ala
1   5

<210> 365
<211> 7
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of artificial sequence: synthetic peptide sequence

<400> 365

Pro Ala Pro Ala Pro Ala Pro
1   5

<210> 366
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 366

Pro Ala Pro Ala Pro Ala Pro Ala
1   5

<210> 367
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 367

Pro Ala Pro Ala Pro Ala Pro Ala Pro
1   5

<210> 368
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 368

Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro
1   5   10   15

<210> 369
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 369

Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala Pro Ala
1 5 10 15

Pro Ala Pro Ala Pro
20

<210> 370
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 370
caguauggac acuguccaaa 20

<210> 371
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 371
ccaccaguau ggacacuguc 20

<210> 372
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 372
caccaguaug gacacugucc 20

<210> 373

<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 373
accaguaugg acacugucca 20

<210> 374
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 374
ccaguaugga cacuguccaa 20

<210> 375
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 375
aguauggaca cuguccaaag 20

<210> 376
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 376
guauggacac uguccaaaga 20

<210> 377
<211> 20
<212> RNA

<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 377
uauggacacu guccaaagag 20

<210> 378
<211> 8811
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 378
atatgccaag tacgcccccc attgacgtca atgacggtaa atggcccgcc tggcattatg 60

cccagtacat gaccttatgg gactttccta cttggcagta catctacgta ttagtcatcg 120

ctattaccat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag cggtttgact 180

cacggggatt tccaagtctc caccccattg acgtcaatgg gagtttgttt tggcaccaaa 240

atcaacggga ctttccaaaa tgtcgtaaca actccgcccc attgacgcaa atgggcggta 300

ggcgtgtacg gtgggaggtc tatataagca gagctggttt agtgaaccgt cagatccgct 360

agagatccgc ggccgctaat acgactcact atagggagag ccgccaccat gaaacggaca 420

gccgacggaa gcgagttcga gtcaccaaag aagaagcgga aagtctctga agtcgagttt 480

agccacgagt attggatgag gcacgcactg accctggcaa agcgagcatg ggatgaaaga 540

gaagtccccg tgggcgccgt gctggtgcac aacaatagag tgatcggaga gggatggaac 600

aggccaatcg gccgccacga ccctaccgca cacgcagaga tcatggcact gaggcaggga 660

ggcctggtca tgcagaatta ccgcctgatc gatgccaccc tgtatgtgac actggagcca 720

tgcgtgatgt gcgcaggagc aatgatccac agcaggatcg gaagagtggt gttcggagca 780

cgggacgcca agaccggcgc agcaggctcc ctgatggatg tgctgcacca ccccggcatg 840

aaccaccggg tggagatcac agagggaatc ctggcagacg agtgcgccgc cctgctgagc 900

gatttcttta gaatgcggag acaggagatc aaggcccaga agaaggcaca gagctccacc 960

gactctggag gatctagcgg aggatcctct ggaagcgaga caccaggcac aagcgagtcc 1020

gccacaccag agagctccgg cggctcctcc ggaggatcct ctgaggtgga gttttcccac 1080

gagtactgga tgagacatgc cctgaccctg gccaagaggg cacgcgatga gagggaggtg 1140

cctgtgggag ccgtgctggt gctgaacaat agagtgatcg gcgagggctg gaacagagcc 1200

atcggcctgc acgacccaac agcccatgcc gaaattatgg ccctgagaca gggcggcctg 1260

gtcatgcaga actacagact gattgacgcc accctgtacg tgacattcga gccttgcgtg 1320

atgtgcgccg gcgccatgat ccactctagg atcggccgcg tggtgtttgg cgtgaggaac 1380

gcaaaaaccg gcgccgcagg ctccctgatg gacgtgctgc actaccccgg catgaatcac 1440

cgcgtcgaaa ttaccgaggg aatcctggca gatgaatgtg ccgccctgct gtgctatttc 1500

tttcggatgc ctagacaggt gttcaatgct cagaagaagg cccagagctc caccgactcc 1560

ggaggatcta gcggaggctc ctctggctct gagacacctg gcacaagcga gagcgcaaca 1620

cctgaaagca gcgggggcag cagcgggggg tcagacaaga agtacagcat cggcctggcc 1680

atcggcacca actctgtggg ctgggccgtg atcaccgacg agtacaaggt gcccagcaag 1740

aaattcaagg tgctgggcaa caccgaccgg cacagcatca agaagaacct gatcggagcc 1800

ctgctgttcg acagcggcga aacagccgag gccacccggc tgaagagaac cgccagaaga 1860

agatacacca gacggaagaa ccggatctgc tatctgcaag agatcttcag caacgagatg 1920

gccaaggtgg acgacagctt cttccacaga ctggaagagt ccttcctggt ggaagaggat 1980

aagaagcacg agcggcaccc catcttcggc aacatcgtgg acgaggtggc ctaccacgag 2040

aagtacccca ccatctacca cctgagaaag aaactggtgg acagcaccga caaggccgac 2100

ctgcggctga tctatctggc cctggcccac atgatcaagt tccggggcca cttcctgatc 2160

gagggcgacc tgaaccccga caacagcgac gtggacaagc tgttcatcca gctggtgcag 2220

acctacaacc agctgttcga ggaaaacccc atcaacgcca gcggcgtgga cgccaaggcc 2280

atcctgtctg ccagactgag caagagcaga cggctggaaa atctgatcgc ccagctgccc 2340

ggcgagaaga agaatggcct gttcggaaac ctgattgccc tgagcctggg cctgaccccc 2400

aacttcaaga gcaacttcga cctggccgag gatgccaaac tgcagctgag caaggacacc 2460

tacgacgacg acctggacaa cctgctggcc cagatcggcg accagtacgc cgacctgttt 2520

ctggccgcca agaacctgtc cgacgccatc ctgctgagcg acatcctgag agtgaacacc 2580

gagatcacca aggcccccct gagcgcctct atgatcaaga gatacgacga gcaccaccag 2640

gacctgaccc tgctgaaagc tctcgtgcgg cagcagctgc ctgagaagta caaagagatt 2700

ttcttcgacc agagcaagaa cggctacgcc ggctacattg acggcggagc cagccaggaa 2760

gagttctaca agttcatcaa gcccatcctg gaaaagatgg acggcaccga ggaactgctc 2820

gtgaagctga acagagagga cctgctgcgg aagcagcgga ccttcgacaa cggcagcatc 2880

ccccaccaga tccacctggg agagctgcac gccattctgc ggcggcagga agatttttac 2940

ccattcctga aggacaaccg ggaaaagatc gagaagatcc tgaccttccg catcccctac 3000

tacgtgggcc ctctggccag gggaaacagc agattcgcct ggatgaccag aaagagcgag 3060

gaaaccatca ccccctggaa cttcgaggaa gtggtggaca agggcgcttc cgcccagagc 3120

ttcatcgagc ggatgaccaa cttcgataag aacctgccca acgagaaggt gctgcccaag 3180

cacagcctgc tgtacgagta cttcaccgtg tataacgagc tgaccaaagt gaaatacgtg 3240

accgagggaa tgagaaagcc cgccttcctg agcggcgagc agaaaaaggc catcgtggac 3300

ctgctgttca agaccaaccg gaaagtgacc gtgaagcagc tgaaagagga ctacttcaag 3360

aaaatcgagt gcttcgactc cgtggaaatc tccggcgtgg aagatcggtt caacgcctcc 3420

ctgggcacat accacgatct gctgaaaatt atcaaggaca aggacttcct ggacaatgag 3480

gaaaacgagg acattctgga agatatcgtg ctgaccctga cactgtttga ggacagagag 3540

atgatcgagg aacggctgaa aacctatgcc cacctgttcg acgacaaagt gatgaagcag 3600

ctgaagcggc ggagatacac cggctggggc aggctgagcc ggaagctgat caacggcatc 3660

cgggacaagc agtccggcaa gacaatcctg gatttcctga agtccgacgg cttcgccaac 3720

agaaacttca tgcagctgat ccacgacgac agcctgacct ttaaagagga catccagaaa 3780

gcccaggtgt ccggccaggg cgatagcctg cacgagcaca ttgccaatct ggccggcagc 3840

cccgccatta agaagggcat cctgcagaca gtgaaggtgg tggacgagct cgtgaaagtg 3900

atgggccggc acaagcccga gaacatcgtg atcgaaatgg ccagagagaa ccagaccacc 3960

cagaagggac agaagaacag ccgcgagaga atgaagcgga tcgaagaggg catcaaagag 4020

ctgggcagcc agatcctgaa agaacacccc gtggaaaaca cccagctgca gaacgagaag 4080

ctgtacctgt actacctgca gaatgggcgg gatatgtacg tggaccagga actggacatc 4140

aaccggctgt ccgactacga tgtggaccat atcgtgcctc agagctttct gaaggacgac 4200

tccatcgaca acaaggtgct gaccagaagc gacaagaacc ggggcaagag cgacaacgtg 4260

ccctccgaag aggtcgtgaa gaagatgaag aactactggc ggcagctgct gaacgccaag 4320

ctgattaccc agagaaagtt cgacaatctg accaaggccg agagaggcgg cctgagcgaa 4380

ctggataagg ccggcttcat caagagacag ctggtggaaa cccggcagat cacaaagcac 4440

gtggcacaga tcctggactc ccggatgaac actaagtacg acgagaatga caagctgatc 4500

cgggaagtga aagtgatcac cctgaagtcc aagctggtgt ccgatttccg gaaggatttc 4560

cagttttaca aagtgcgcga gatcaacaac taccaccacg cccacgacgc ctacctgaac 4620

gccgtcgtgg gaaccgccct gatcaaaaag taccctaagc tggaaagcga gttcgtgtac 4680

ggcgactaca aggtgtacga cgtgcggaag atgatcgcca agagcgagca ggaaatcggc 4740

aaggctaccg ccaagtactt cttctacagc aacatcatga acttttcaa gaccgagatt 4800

accctggcca acggcgagat ccggaagcgg cctctgatcg agacaaacgg cgaaaccggg 4860

gagatcgtgt gggataaggg ccgggatttt gccaccgtgc ggaaagtgct gagcatgccc 4920

caagtgaata tcgtgaaaaa gaccgaggtg cagacaggcg gcttcagcaa agagtctatc 4980

ctgcccaaga ggaacagcga taagctgatc gccagaaaga aggactggga ccctaagaag 5040

tacggcggct tcgacagccc caccgtggcc tattctgtgc tggtggtggc caaagtggaa 5100

aagggcaagt ccaagaaact gaagagtgtg aaagagctgc tggggatcac catcatggaa 5160

agaagcagct tcgagaagaa tcccatcgac tttctggaag ccaagggcta caaagaagtg 5220

aaaaaggacc tgatcatcaa gctgcctaag tactccctgt tcgagctgga aaacggccgg 5280

aagagaatgc tggcctctgc cggcgaactg cagaagggaa acgaactggc cctgccctcc 5340

aaatatgtga acttcctgta cctggccagc cactatgaga agctgaaggg ctcccccgag 5400

gataatgagc agaaacagct gtttgtggaa cagcacaagc actacctgga cgagatcatc 5460

gagcagatca gcgagttctc caagagagtg atcctggccg acgctaatct ggacaaagtg 5520

ctgtccgcct acaacaagca ccgggataag cccatcagag agcaggccga gaatatcatc 5580

cacctgttta ccctgaccaa tctgggagcc cctgccgcct tcaagtactt tgacaccacc 5640

atcgaccgga agaggtacac cagcaccaaa gaggtgctgg acgccaccct gatccaccag 5700

agcatcaccg gcctgtacga gacacggatc gacctgtctc agctgggagg tgactctggc 5760

ggctcaaaaa gaaccgccga cggcagcgaa ttcgagccca agaagaagag gaaagtctaa 5820

ccggtcatca tcaccatcac cattgagttt aaacccgctg atcagcctcg actgtgcctt 5880

ctagttgcca gccatctgtt gtttgcccct cccccgtgcc ttccttgacc ctggaaggtg 5940

ccactcccac tgtcctttcc taataaaatg aggaaattgc atcgcattgt ctgagtaggt 6000

gtcattctat tctgggggggt ggggtggggc aggacagcaa gggggaggat tgggaagaca 6060

atagcaggca tgctggggat gcggtgggct ctatggcttc tgaggcggaa agaaccagct 6120

ggggctcgat accgtcgacc tctagctaga gcttggcgta atcatggtca tagctgtttc 6180

ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat acgagccgga agcataaagt 6240

gtaaagccta gggtgcctaa tgagtgagct aactcacatt aattgcgttg cgctcactgc 6300

ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc caacgcgcgg 6360

ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc gctcactgac tcgctgcgct 6420

cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa ggcggtaata cggttatcca 6480

cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga 6540

accgtaaaaa ggccgcgttg ctggcgtttt tccataggct ccgcccccct gacgagcatc 6600

acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg 6660

cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat 6720

acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt 6780

atctcagttc ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa ccccccgttc 6840

agcccgaccg ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg 6900

acttatcgcc actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg 6960

gtgctacaga gttcttgaag tggtggccta actacggcta cactagaaga acagtatttg 7020

gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg 7080

gcaaacaaac caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca 7140

gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac actcagtgga 7200

acgaaaactc acgttaaggg attttggtca tgagattatc aaaaaggatc ttcacctaga 7260

tccttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggt 7320

ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcgtt 7380

catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat 7440

ctggccccag tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag 7500

caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgcct 7560

ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt 7620

tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg 7680

cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatccccc atgttgtgca 7740

aaaaagcggt tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt 7800

tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca tccgtaagat 7860

gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac 7920

cgagttgctc ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa 7980

aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt 8040

tgagatccag ttcgatgtaa cccactcgtg cacccaactg atcttcagca tcttttactt 8100

tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa 8160

gggcgacacg gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt 8220

atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa 8280

taggggttcc gcgcacattt ccccgaaaag tgccacctga cgtcgacgga tcgggagatc 8340

gatctcccga tcccctaggg tcgactctca gtacaatctg ctctgatgcc gcatagttaa 8400

gccagtatct gctccctgct tgtgtgttgg aggtcgctga gtagtgcgcg agcaaaattt 8460

aagctacaac aaggcaaggc ttgaccgaca attgcatgaa gaatctgctt agggttaggc 8520

gttttgcgct gcttcgcgat gtacgggcca gatatacgcg ttgacattga ttattgacta 8580

gttattaata gtaatcaatt acggggtcat tagttcatag cccatatatg gagttccgcg 8640

ttacataact tacggtaaat ggcccgcctg gctgaccgcc caacgacccc cgcccattga 8700

cgtcaataat gacgtatgtt cccatagtaa cgccaatagg gactttccat tgacgtcaat 8760

gggtggagta tttacggtaa actgcccact tggcagtaca tcaagtgtat c 8811

<210> 379
<211> 357
<212> PRT
<213> Homo sapiens

<400> 379

Met Glu Glu Gly Met Asn Val Leu His Asp Phe Gly Ile Gln Ser Thr
1 5 10 15

```
His Tyr Leu Gln Val Asn Tyr Gln Asp Ser Gln Asp Trp Phe Ile Leu
            20                  25                  30

Val Ser Val Ile Ala Asp Leu Arg Asn Ala Phe Tyr Val Leu Phe Pro
        35                  40                  45

Ile Trp Phe His Leu Gln Glu Ala Val Gly Ile Lys Leu Leu Trp Val
    50                  55                  60

Ala Val Ile Gly Asp Trp Leu Asn Leu Val Phe Lys Trp Ile Leu Phe
65                  70                  75                  80

Gly Gln Arg Pro Tyr Trp Trp Val Leu Asp Thr Asp Tyr Tyr Ser Asn
                85                  90                  95

Thr Ser Val Pro Leu Ile Lys Gln Phe Pro Val Thr Cys Glu Thr Gly
            100                 105                 110

Pro Gly Ser Pro Ser Gly His Ala Met Gly Thr Ala Gly Val Tyr Tyr
        115                 120                 125

Val Met Val Thr Ser Thr Leu Ser Ile Phe Gln Gly Lys Ile Lys Pro
    130                 135                 140

Thr Tyr Arg Phe Arg Cys Leu Asn Val Ile Leu Trp Leu Gly Phe Trp
145                 150                 155                 160

Ala Val Gln Leu Asn Val Cys Leu Ser Arg Ile Tyr Leu Ala Ala His
                165                 170                 175

Phe Pro His Gln Val Val Ala Gly Val Leu Ser Gly Ile Ala Val Ala
            180                 185                 190

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

Tyr Phe Leu Ile Thr Phe Phe Leu Phe Ser Phe Ala Ile Gly Phe Tyr
    210                 215                 220
```

```
Leu Leu Leu Lys Gly Leu Gly Val Asp Leu Leu Trp Thr Leu Glu Lys
225                 230                 235                 240

Ala Gln Arg Trp Cys Glu Gln Pro Glu Trp Val His Ile Asp Thr Thr
                245                 250                 255

Pro Phe Ala Ser Leu Leu Lys Asn Leu Gly Thr Leu Phe Gly Leu Gly
            260                 265                 270

Leu Ala Leu Asn Ser Ser Met Tyr Arg Glu Ser Cys Lys Gly Lys Leu
        275                 280                 285

Ser Lys Trp Leu Pro Phe Arg Leu Ser Ser Ile Val Ala Ser Leu Val
    290                 295                 300

Leu Leu His Val Phe Asp Ser Leu Lys Pro Pro Ser Gln Val Glu Leu
305                 310                 315                 320

Val Phe Tyr Val Leu Ser Phe Cys Lys Ser Ala Val Val Pro Leu Ala
                325                 330                 335

Ser Val Ser Val Ile Pro Tyr Cys Leu Ala Gln Val Leu Gly Gln Pro
            340                 345                 350

His Lys Lys Ser Leu
        355
```

<210> 380
<211> 13637
<212> DNA
<213> Homo sapiens

<400> 380

```
atagcagagc aatcaccacc aagcctggaa taactgcaag ggctctgctg acatcttcct       60

gaggtgccaa ggaaatgagg atggaggaag gaatgaatgt tctccatgac tttgggatcc      120

agtcaacaca ttacctccag gtgaattacc aagactccca ggactggttc atcttggtgt      180

ccgtgatcgc agacctcagg aatgccttct acgtcctctt ccccatctgg ttccatcttc      240

aggaagctgt gggcattaaa ctcctttggg tagctgtgat tggagactgg ctcaacctcg      300

tctttaagtg gtaagaacca tatagagagg agatcagcaa gaaaagaggc tggcattcgc      360
```

tctcgcaatg tctgtccatc agaagttgct ttccccaggc tattcaggaa gccacgggct 420

actcatgctt ccaacccctc tctctgactt tggatcatct acataaaggg ggaagacaga 480

aaaaatccta ccagtgagtt gaaaatacag gaaagcctat ttcatatggg ttaaagggta 540

ggacagttga atttcgtgaa aagtctgagt tatataggct ttgagcaaag agttttatta 600

gtatgaagca gaagaggtaa cataaagaaa gatgtatggg gccaggcatg gtggctcaca 660

cctgtaatcc cagcactttg ggaggccgag gtgggcgaat cactcctggg tgaactcagg 720

agttcaagac cagcctgggc aacatggcga aactccatct ctacaaaaac attacgaaaa 780

ttagctgggc gtgttggtgc tgtagtccca gctactcagg aggctgaggt gagaggcgga 840

ggaggttgca gtgagtcaag atcatgccac tgcactccag cctgggcaac agagtaagac 900

cctgtctcaa aaaaaaaaaa aagatagatg atgtatgctg tatgaaaaaa ggaaacacac 960

agatgattca acagcctgtt ttgtggggta atgaaaagtc accctgggaa ctgggctcca 1020

gccctcgttc tgccacccac caactacatg tccttggcaa gtcatatcaa ttatctgagt 1080

ttctgtttta taatctacaa ataggttatc tctggcagct taataataat cagggttaac 1140

atttattaaa cagtgtgtgc cagtccatgt gctatgtgct tttctgtgag gtagttactg 1200

ctatttacag aaacagtaga tgcagagacc aaggtgctga gttaaatgat taggccaaca 1260

aggttagtac atgccgagcc aggatggaag cccaggtagg caggctggct tccgcggcaa 1320

tgctcttatg aactatgtta cgtccagtgc tgataaactg actctctggg gagcagggga 1380

aagccctgag tttagcattt gccaatttct atcacgtaaa cattcccatt ctggccactt 1440

tctttctttc tttcttttgt ttgtttgttt gagatggagt ctcgcactgt tgcctggctg 1500

gagtgcaatg gtgcaatctc agctcactgc aacctctgcc tctccggttc aagtgattct 1560

cctgcctcag cctcccaagt agctgggatt acaggtgccc gccaccatgc ccagctaatt 1620

ttttttgtat ttttagtaga gacatggttt cactatgttg actaggctgg tctcgaactc 1680

ctgacctcat gatctgcctg ccttggcctc cctaagtgct aggattacag gcgtgagcca 1740

ctacacccag ccgcatgatt ctaaaaaata aaaagatgaa gtgttattcc aaacatctga 1800

tctccattga agaaccatgc aatctctctg ggttgataga ggccagagtt agtggctctc 1860

cctgatttcg gtgagaaatc actattccac catcacggga taaaaggcat cctgactggc 1920

ggttgacacc tatttccaca gtgaaagata tatctagtac ttttaaaggg gaagtggttt 1980

gtctgagata ctctgtttca aagtagagag gatacagaac aagcatctga agctatatac 2040

atccttacag agagcaattc tgatggaaat gcaggccatg tttccctggg gggggctcgt 2100

cctaggggct ggagtgcatt ctctgatgtc agaggaaatg caagattccc tgaggcctga 2160

gggaacccat ggtatatgca agtccaagtt tcaaactgta gttccatatg cattcttcca 2220

ggacaaatac ttcttgaggt taaaaaaaaa aagtcacata gctgccattt tatggatttc 2280

aggatttttt ttttttttt tttgagatgg agtcttgctc tgtcacccag cctgtagtgc 2340

agtggcataa tctcggctca cggcaacctc cgcctcccag gttcaagcga ttctcttgcc 2400

ttagcctccc gagtagctgg gattacagtc acgcaccacc acatctggct aattctttat 2460

atttttggt agaaacggtg tttcaccatg ttggccaggc tggtctcaaa ctcctgacct 2520

catgtgatct gcctgccttg gcctcccaaa gtgctgagat tacaggtgtg agccaccgcg 2580

cctgcctgga gttcagaatc ttgggcttca ttatttgtgt ttaaatagat catacagtca 2640

ggcacggtgg ctcatgcctg taatcccagc actttgggag gctgaggtgg gaggattgcc 2700

tgagttcagg agatggagac cagcctgggc aacatggtga aaccccgtct ctactaaaaa 2760

tacaaaaact agctggatgt ggtggcacac acctgtagtc ccagctattc aggaggctga 2820

ggtgggagga tcccaggagg tagaggtcac aatgagccga gattgcgcca ctgcactcca 2880

ggctgggtta ctgagccaga tcctgtctca aaaaaaaaaa agataataca ttcaaacagt 2940

tcaaaatgca aaagttacat acataaggaa gtgtcatgaa atatctccct ctcacacttc 3000

tccccagcca cccagttctc ccttctagag gcaacatgtg aaatccttct caggctacac 3060

tcttcttgaa ggtgtaggct ttgggcaaaa gcattcattc agtaacccca gaaacttgtt 3120

ctgtttttcc ataggattct ctttggacag cgtccatact ggtgggtttt ggatactgac 3180

tactacagca acacttccgt gcccctgata aagcagttcc ctgtaacctg tgagactgga 3240

ccaggtaagc gtcccagccc ctgcagacag aagctgagtg gacctcgttt acctgttatg 3300

gatgaaactg accttgaggg gacatgagga gagccattcc tttgtacttt tgtcatgctc 3360

ttcaattggc acaaattaat tcacttctgc aatactttcc tgaatagcac agtagtattg 3420

gaaatctgcc tattacagaa cctggatgga gtccagagag gcacgggcat ccatgggcaa 3480

agggctcgtg agagtcaccg ccctgcagcg ctgtgtcctg agaaaggagg gggcagaagc 3540

ctgagcttct gggggtcctt cccaatggcc tggcccactg gatgtgccct cctgagctga 3600

ccgtccaatc ccttgccctc tctgtgccta cgttttatta gttacagcca gatggttact 3660

gtcaaatcaa atgatagatt tcattttcag tatgtaatag gaagcccctc cctcacccta 3720

aagtctcagc tgccctctaa gactagtact ctctaaggta ctagtatccc ttcctcagag 3780

acccttttccc tgaccccaaa actagggaag gtcccttagt tatttgctct cacagaccac 3840

gcatttacct cagagcatat tcactcattc agctgttact taccaagcac ctactgggag 3900

ctatacactg ttctatgtgc tagggatacc tctgtcagtg aacaacacag acacaaagat 3960

ccctgccctt gtggagctga aatctgaata gaggaggtga aatatacaaa aattataata 4020

aataagtaaa ctaggccagt tgtggttgct catgcctgta atcccagcac tttgggaagc 4080

caaggtaggt agatcacctg aggtcaggag ttcaaaacca gcctggccaa cattgcaaaa 4140

tcctgtcttt actaaaaatg gaaaaattgg tcaggcgtga tggcacacgc ctgtagtctc 4200

agctacctgg gaggctgagg caggagaatc gcttgaacct gggaggcaga ggttgcagtg 4260

aaccgagatc ggaccactgc actccagcct gaatgacaga acgagactct gtctcaaaaa 4320

aaaagtaaac tattaatatg taggataggc caggcacggt ggctcaccct gtaatcccag 4380

cactttggga ggctgaggcg ggtggatcac ctgaggtgag gagttcaaga ccagcctggc 4440

caacatggca aaaccctgtc tctactaaaa atacaaaaat tagctgggtg tcctggtgca 4500

tgcctgtaat ctgagctact caggaggcta aggcaggaga atcgcttgaa cctgggaggt 4560

ggtgagccaa gattgcgcca ttgcactcca gcctgggcga caaaatgaga caccatctga 4620

aaaaaaaaaa aaaatatata tatatataca cacacacaca cacacacaca cacacacaca 4680

tataatacta gaaaatgatt gtttataggc aaaaaaaaaa aaaaagaaga agaagaagaa 4740

aaggaaagga gaaggaaaga aggaccaaac atcttttgta gaaatatgtt tgctttcatc 4800

ataacagctt gttatcaagg atgaatttct ccctgaaatt aatggaggca cagactggaa 4860

agtttaaagt ggctttaaga ggttatttta tttagtcctc tgtcttaata gaagcaaatt 4920

attatctctg ctccttaggt agagtagcta aggctcagaa agtaggccgg gcgcggtggc 4980

tcacgcctgt aatcctagca ctttgggagg ccaacgcagg tggatcacct gaggtcagga 5040

gtttgagacc agcctggcca acatggtgaa acctcgtcac taataaaaaa atacaaaaac 5100

ttagccaggc atggtggcgg gcgcctgtaa tcccagctac ccaggaggct gcggcaggag 5160

aatcacttca acccgggagg cagaggttgc agtgagctga aatcacacca ctgcactcca 5220

gccttggtga cagagaaaga ttctgtcagg aaaaaaaaaa aaaagtttaa atgaattacc 5280

caaggtatat aattgttagt gttagaagga agaagaaggg agggaggaag gaagggagaa 5340

agaaagggaa ggaggaaggg agggagggaa gaaagccttt atttatctat ggggttccct 5400

ggaaagcagg ctgaaatgga gattcacgtg caggagttta gatactctgg ggaactatac 5460

ttgtagaagg gaaggaacag gaacagggca gaaggagagg tccggttgtg attctgcctc 5520

atccaacccc acagcgagct ctgaagctgg ggatggctcc tcagagttgg tccaagttgg 5580

gacaagggaa tcagaccctg gggagagcgt aaccttgatc aaggcgactc tctttagccc 5640

agggcaatgc caggagaagg ctgagagcag aaagccatct accatcacac tctcaacagc 5700

tacgaaataa gtcctgcagt tcaggaggga ggtctgggcg gcacatctca ggaccctcta 5760

tctctcaggg tagaggaatt aagaatggga tgggaaccag acgggccatg gtggctcaca 5820

cctataatcc caacactttg ggaggccaag ggtaggagga ttgcttgagc ccaagagttc 5880

aaaaccagcc tgggcaaaaa caatcaaaca aacaaacaaa acacatttaa aaaatttgct 5940

gtgtgtggtg gtgtgcacct gtggtcccag ctactcaggg ggctgaggtg ggaggattgc 6000

ttgagtccag gaggtcgagg ctgcagtgag ctatgatcat ggcactgcat tgcagcctag 6060

gagacaaagc aagacactgt ctctaaaaaa acaaaaaaca aacaaataaa aaaacggaac 6120

cggttgcaag cagggttaaa tagcgtggtc agagtaggac tcactgagaa tatgagatct 6180

gagtcaagtc ttcaaggatg tgaggaagta agtttctggc agaagagctg tgaagggctg 6240

tctggccaga gaagattgca atgcaaaagc cctgaggtgg gaacgtgttt ggtgtgttta 6300

aaggaaagca atgaggccag tgtagccaga acagagtgtg caaggagaga aggaacagaa 6360

gatgtggagg gcagatcagt ttgtaattgt acgcccagta tgctgattct ttgtgtaatc 6420

tccagactgt attaaactgc aagagcaggg cccctctctg gctttgctca tcattgtatt 6480

cccagagcct tgcacaatgc ttggtgcata ggagatggaa atttgttaaa taaatgaatt 6540

atggataacg aatggatggt aagatgggtg gatggatggg gggtgaacgg atggatgggg 6600

ggtgaatgga tggatgaatg ggtagatggg tggatagggg gatggctggg tggctgggta 6660

gatgatgcac tgtctcccag atgaggacct tttcaccttt actccattct ctttcctgcc 6720

ctttagggag cccctctggc catgccatgg gcacagcagg tgtatactac gtgatggtca 6780

catctactct ttccatcttt cagggaaaga taaagccgac ctacagattt cggtaagaac 6840

tcaccactgg ggtgtaggtg gtggagggca ggaggcagct ctctctgtag ctgacacacc 6900

acgtattctt cctcacatcc ccctagcccg ctcccacacc tgggcagccg ctgattaaga 6960

gttgtggcac tttggatagg gataaacctc agagtcaggg aatgtttggg ctgaaaggga 7020

tccagtagtg caatccgttg ttttacagat aaggaaacaa agcccaacac catgaaggga 7080

cttataaaaa taaggtagtg aagtagcagc agggcttaaa taaaaaccca tgtctgtacc 7140

aaccacagag tcacccatcc aggttaaaat aaccagagaa acagaagata ttcctactac 7200

agagaattcc gggtgtgcag ccacagtgca aatccttttt atttttattt ttgagatgca 7260

gtctcgctct gtcatccagg ctgaagtgca gtggcacgat catgtctcgc tgcaacctct 7320

gcctcccagg ctcaagcgat cctcccacct cagccatctg agtagctggg accacaggcc 7380

acacaccaca cccagctaat ttctcgtatc tttttgtaga gacagagttc tgctatgttg 7440

cccaggctca ggctggtctt gatctcaagc aattggcttg cctcagcctc ctaaaatatt 7500

gggattacag gcatgagcca ccgcgccagc catgcaaatc cttaattatc aaacagataa 7560

aatagggaag ttaaaattca tatacacaag ggttaaccac ttgccacagg catttttttt 7620

tttttttga gacggaatct cgctctgttg cccaggctgg agtgcagtgg cgccatctcg 7680

cctcactgca acctccgctt cctgggttca agctattctt ctgcctcagc ctaccgagta 7740

gctgggacta caggcacgtg ccaccacacc tggctaattt ttttatttt agtagagatg 7800

gggtttcacc atattggcca ggctggtctt gaactcctga cctagtgatc catccgcctc 7860

agcctcccaa agtgctggga ttgcaggcat gagccaccgc gcctggcctt ttttttttt 7920

ttttgagacg gagttttgct cttgttgccc aggctagagt gcagtggcgc agtctcggct 7980

cactgtaacc tccacctcct gagttcaagc aattctcctg cctcagcctc tcaaatagct 8040

gggattacag gcgtgagcca ccccacctgg ctaattttgt aattttttt ttagtagaga 8100

tggggtttca cctgttgatc aggctggtct caaactcctg acctcaagtg atccacccac 8160

ctcggcctcc caaagtgctg ggattacaag cataagccac cgtgcctggt caattttgat 8220

ctttttaaa gagacagggg tcttgctatg ttgcccagac tagtcttgaa ctcctggcct 8280

caagtgatcc tctcacctcg gcctcccaaa gtattgggat tacaggtctg agccgctgca 8340

cccagccccc aacaggcatc tttggacttt tgagtactgg ctttaattta caaaaattcc 8400

actgagagca cctaagtttg ccaggctcca acatttctgc aggggctgtt ttctttgctg 8460

aaggatctgc acctgtgttc tgttatggtt gcctcttctg ttgcaggtgc ttgaatgtca 8520

ttttgtggtt gggattctgg gctgtgcagc tgaatgtctg tctgtcacga atctaccttg 8580

ctgctcattt tcctcatcaa gttgttgctg gagtcctgtc aggtatgggc tgatctgact 8640

cccttccttc tcccccaaac cccattccgt ttctctccct aatcaggaca aaatcccagc 8700

attccagcca catcctgtgt gtaatcagta ctgttagcat ttctgtgggt tgaaagtcaa 8760

gaatgagcaa cttgaaatga ttaatttcta taagagtgcc cagatctata gaatgaattg 8820

tgtagaagtt accatacatc aaattaacgc accaaattga attagcttga aatctcagag 8880

ctttttacaa tctttatttc ttactggtct tcaacaggcc ctaatttact tttcagggaa 8940

tctgccaaat ttaacaaatt aacacgatgt cctaggaaag ctgttcattt aaatacattc 9000

atttgcaaac ctaatagata actgcagttg atctctttta taggttcaga gttttgaata 9060

tgttttttt tgttttttt ttttgagatg gagtctcgct ctgtgaccca ggctagagtg 9120

cagtggtgcg atctcggctc actgcaagct ccacctcctg ggttcacgcc attctcctgc 9180

ctcagcctct ccgagtagct gggactacag gcgcccgcca ccatgcccgg ctaatttttt 9240

gtatttttag cagagacggg gtttcaccgt ggtcttgatc tcctgacctc gtgatccgcc 9300

cgcctcggcc tcccaaagcg ctgggattac aagggtgagc caccgcaccc tgcctgaata 9360

tgtgttttct tagatccaat taacaagggt aagacaagat ttaagttaag cataagaaag 9420

attttgtggg aggcactgga atataagacc ttaacaaaac tgtggaattt ctcccctgga 9480

gatttgtaag aacggaacat agcagcattc aaagaagaat gttgagaaca agggagataa 9540

tggtttcatg gtaatcacaa aagtaacaca gcatttagta ctgggttcca tgtttgagga 9600

agaacctgga agccatatca catgaaaaac ctgggaatgt ttaggttaga gagaataact 9660

gtgttcaaat gtgtgacaga gggactagat tcatcactta ctaactcctg cagaaagaac 9720

tgagaaaaat agacagtatt agaggggggac cagtttcaca cagacaagga agaactattc 9780

agcaatcaat tccgttcaaa gataaaatgg actgttatag tgggggtgag ctccctacct 9840

ctgagggtat ttcaagtaga gataggagga cctcctggta ggaaatttgc atacggtggg 9900

agattgtacg tgatatggca cctccatctg aaagagtcta tattgagggc aggctggagt 9960

cacacatggg aataagccag gcgaccctcc catctgccat ctgtgattta attccacagt 10020

cgcagaacgg atggcatgtc acccactcct ccaaacccac ctctagcaaa ggtcccaaat 10080

ccttcctatc tctcacagtc atgctttctt ccactcaggc attgctgtta cagaaacttt 10140

cagccacatc cacagcatct ataatgccag cctcaagaaa tattttctca ttaccttctt 10200

cctgttcagc ttcgccatcg gattttatct gctgctcaag ggactgggtg tagacctcct 10260

gtggactctg gagaaagccc agaggtggtg cgagcagcca gaatgggtcc acattgacac 10320

cacacccttt gccagcctcc tcaagaacct gggcacgctc tttggcctgg ggctggctct 10380

caactccagc atgtacaggg agagctgcaa ggggaaactc agcaagtggc tcccattccg 10440

cctcagctct attgtagcct ccctcgtcct cctgcacgtc tttgactcct tgaaaccccc 10500

atcccaagtc gagctggtct tctacgtctt gtccttctgc aagagtgcgg tagtgcccct 10560

ggcatccgtc agtgtcatcc cctactgcct cgcccaggtc ctgggccagc cgcacaagaa 10620

gtcgttgtaa gagatgtgga gtcttcggtg tttaaagtca acaaccatgc cagggattga 10680

ggaggactac tatttgaagc aatgggcact ggtatttgga gcaagtgaca tgccatccat 10740

tctgccgtcg tggaattaaa tcacggatgg cagattggag ggtcgcctgg cttattccca 10800

tgtgtgactc cagcctgccc tcagcacaga ctctttcaga tggaggtgcc atatcacgta 10860

caccatatgc aagtttcccg ccaggaggtc ctcctctctc tacttgaata ctctcacaag 10920

tagggagctc actcccactg gaacagccca ttttatcttt gaatggtctt ctgccagccc 10980

attttgaggc cagaggtgct gtcagctcag gtggtcctct tttacaatcc taatcatatt 11040

gggtaatgtt tttgaaaagc taatgaagct attgagaaag acctgttgct agaagttggg 11100

ttgttctgga ttttcccctg aagacttact tattcttccg tcacatatac aaaagcaaga 11160

cttccaggta gggccagctc acaagcccag gctggagatc ctaactgaga attttctacc 11220

tgtgttcatt cttaccgaga aaaggagaaa ggagctctga atctgatagg aaaagaaggc 11280

tgcctaagga ggagttttta gtatgtggcg tatcatgcaa gtgctatgcc aagccatgtc 11340

taaatggctt taattatata gtaatgcact ctcagtaatg ggggaccagc ttaagtataa 11400

ttaatagatg gttagtgggg taattctgct tctagtattt tttttactgt gcatacatgt 11460

tcatcgtatt tccttggatt tctgaatggc tgcagtgacc cagatattgc actaggtcaa 11520

aacattcagg tatagctgac atctcctcta tcacattaca tcatcctcct tataagccca 11580

gctctgcttt ttccagattc ttccactggc tccacatcca ccccactgga tcttcagaag 11640

gctagagggc gactctggtg gtgcttttgt atgtttcaat taggctctga aatcttgggc 11700

aaaatgacaa ggggagggcc aggattcctc tctcaggtca ctccagtgtt acttttaatt 11760

cctagagggt aaatatgact cctttctcta tcccaagcca accaagagca cattcttaaa 11820

ggaaaagtca acatcttctc tctttttttt ttttttgag acagggtctc actatgttgc 11880

ccaggctgct cttgaattcc tgggctcaag cagtcctccc accctaccac agcgtcccgc 11940

gtagctggga ctacaggtgc aagccactat gtccagctag ccaactcctc cttgcctgct 12000

tttctttttt tttctttttt tgagacggcg cacctatcac ccaggctgga gtggagtggc 12060

acgatcttgg ctcactgcaa cctcttcctc ctggttcaag cgattctcat gtctcagcct 12120

cctcagtagc taggactacc ggcgtgcacc accatgccag gctaattttt atatttttag 12180

aattttagaa gagatgggat ttcatcatgt tggccaggct ggtctcgaac tcctgacctc 12240

aagtgatcca cctgccttgg cctcccaagg tgctaggatt acaggcatga gccaccgcac 12300

cgggccctcc ttgcctgttt ttcaatctca tctgatatgc agagtatttc tgccccaccc 12360

acctaccccc caaaaaaagc tgaagcctat ttatttgaaa gtccttgttt ttgctactaa 12420

ttatatagta taccatacat tatcattcaa aacaaccatc ctgctcataa catctttgaa 12480

aagaaaaata tatatgtgca gtattttatt aaagcaacat tttatttaag aataaagtct 12540

tgttaattac tatattttag atgcaatgtg atctgaagtt tctaattctg gcccaactaa 12600

atttctagct ctgtttccct aaacaaataa tttggtttct ctgtgcctgc attttccctt 12660

tggagaagaa aagtgctctc tcttgagttg accgagagtc ccattaggga tagggagact 12720

taaatgcatc cacaggggca caggcagagt tgagcacata aacggaggcc caaaatcagc 12780

atagaaccag aaagattcag agttggccaa gaatgaacat tggctaccag accacaagtc 12840

agcatgagtt gctctatggc atcaaattgc aacttgagag tagatgggca gggtcactat 12900

caaattaagc aatcagggca cacaagttgc agtaacacaa caagactagg ccagctctgg 12960

aatccagtaa ctcagtgtca gcaaggtttt gggttatagt tcaagaaagt ctaaacagag 13020

ccagtcacag caccaaggaa tgctcaaggg agctattgca ggtttctctg ctaagagatt 13080

tatttcatcc tgggtgcagg gttcgacctc caaaggcctc aaatcatcac cgtatcaatg 13140

gatttcctga gggtaagctc cgctatttca cacctgaact ccggagtctg tatattcagg 13200

gaagattgca ttctcctact ggatttgggc tctcagaggg cgttgtggga accaggcccc 13260

tcacagaatc aaatggtccc aaccagggag aaagaaaata gtctttttt ttttttaat 13320

agagatgggg gtctcactat gctgcccagg ctggtcttga actcctgggt tcaagtgatc 13380

ctcctgcctc agcctcccaa agtgctggga ttacagtgtg agccactgcg cttggccaga 13440

aatggttttg atctgtctga actgaaccct actgcttagg catagcccca tccttgataa 13500

tctatttgct cccaaggacc aagtccaaga tccttacaag aaaggtctgc cagaaagtaa 13560

atactgcccc cactccctga agtttatgag gttgataaga aaacataaca gataaagttt 13620

attgagtgct aacttta 13637

<210> 381
<211> 306
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 381
tgcctgtcat acgaaaccga gatactgaca gtagaatatg gccttctgcc aatcgggaag 60

attgtggaga aacggataga atgcacagtt tactctgtcg ataacaatgg taacatttat 120

actcagccag ttgcccagtg gcacgaccgg ggagagcagg aagtattcga atactgtctg 180

gaggatggaa gtctcattag ggccactaag gaccacaaat ttatgacagt cgatggccag 240

atgctgccta tagacgaaat ctttgagcga gagttggacc tcatgcgagt tgacaacctt 300

cctaat 306

<210> 382
<211> 102
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 382

Cys Leu Ser Tyr Glu Thr Glu Ile Leu Thr Val Glu Tyr Gly Leu Leu
1 5 10 15

Pro Ile Gly Lys Ile Val Glu Lys Arg Ile Glu Cys Thr Val Tyr Ser
20 25 30

Val Asp Asn Asn Gly Asn Ile Tyr Thr Gln Pro Val Ala Gln Trp His
35 40 45

Asp Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
50 55 60

Leu Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Val Asp Gly Gln
65 70 75 80

Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Glu Leu Asp Leu Met Arg
85 90 95

Val Asp Asn Leu Pro Asn
100

<210> 383
<211> 108
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 383
atgatcaaga tagctacaag gaagtatctt ggcaaacaaa acgtttatga tattggagtc 60

gaaagagatc acaactttgc tctgaagaac ggattcatag cttctaat 108

<210> 384
<211> 36
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 384

Met Ile Lys Ile Ala Thr Arg Lys Tyr Leu Gly Lys Gln Asn Val Tyr
1 5 10 15

Asp Ile Gly Val Glu Arg Asp His Asn Phe Ala Leu Lys Asn Gly Phe
20 25 30

Ile Ala Ser Asn
35

<210> 385
<211> 303
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 385
tgcctgtctt atgataccga gatacttacc gttgaatatg gcttcttgcc tattggaaag 60

attgtcgaag agagaattga atgcacagta tatactgtag acaagaatgg tttcgtttac 120

acacagccca ttgctcaatg gcacaatcgc ggcgaacaag aagtatttga gtactgtctc 180

gaggatggaa gcatcatacg agcaactaaa gatcataaat tcatgaccac tgacgggcag 240

atgttgccaa tagatgagat attcgagcgg ggcttggatc tcaaacaagt ggatggattg 300

cca 303

<210> 386
<211> 101
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 386

Cys Leu Ser Tyr Asp Thr Glu Ile Leu Thr Val Glu Tyr Gly Phe Leu
1 5 10 15

Pro Ile Gly Lys Ile Val Glu Glu Arg Ile Glu Cys Thr Val Tyr Thr
20 25 30

Val Asp Lys Asn Gly Phe Val Tyr Thr Gln Pro Ile Ala Gln Trp His
35 40 45

Asn Arg Gly Glu Gln Glu Val Phe Glu Tyr Cys Leu Glu Asp Gly Ser
50 55 60

```
Ile Ile Arg Ala Thr Lys Asp His Lys Phe Met Thr Thr Asp Gly Gln
65                  70                  75                  80


Met Leu Pro Ile Asp Glu Ile Phe Glu Arg Gly Leu Asp Leu Lys Gln
                85                  90                  95


Val Asp Gly Leu Pro
            100


<210>  387
<211>  159
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: synthetic polypeptide
       sequence

<400>  387
atgaagagga ctgccgatgg atcagagttt gaatctccca agaagaagag gaaagtaaag       60

ataatatctc gaaaaagtct tggtacccaa aatgtctatg atattggagt ggagaaagat      120

cacaacttcc ttctcaagaa cggtctcgta gccagcaac                             159


<210>  388
<211>  53
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Description of artificial sequence: synthetic polypeptide
       sequence

<400>  388

Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                   10                  15


Arg Lys Val Lys Ile Ile Ser Arg Lys Ser Leu Gly Thr Gln Asn Val
            20                  25                  30


Tyr Asp Ile Gly Val Glu Lys Asp His Asn Phe Leu Leu Lys Asn Gly
        35                  40                  45


Leu Val Ala Ser Asn
```

50

<210> 389
<211> 125
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 389
guuuuuguac ucucaagauu uaaguaacug uacaacgaaa cuuacacagu uacuuaaauc 60

uugcagaagc uacaaagaua aggcuucaug ccgaaaucaa cacccuguca uuuuauggca 120

gggug 125

<210> 390
<211> 80
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 390
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80

<210> 391
<211> 167
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 391

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 Trp Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

```
Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
        35                  40                  45

Ile Gly Arg 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 Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp 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 Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp
                165
```

<210> 392
<211> 178
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 392

```
Met Arg Arg Ala Phe Ile Thr Gly Val Phe Phe Leu Ser Glu Val Glu
1               5                   10                  15

Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg
            20                  25                  30
```

```
Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val His Asn
        35                  40                  45

Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro Ile Gly Arg His Asp
    50                  55                  60

Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
65                  70                  75                  80

Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Leu Glu
                85                  90                  95

Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg
            100                 105                 110

Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly Ala Ala Gly Ser Leu
        115                 120                 125

Met Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr
    130                 135                 140

Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Ser Asp Phe Phe
145                 150                 155                 160

Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys Lys Ala Gln Ser Ser
                165                 170                 175

Thr Asp
```

<210> 393
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 393
gggcattaaa ctcctttggg 20

<210> 394
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 394
agtctcacag gttacaggga 20

<210> 395
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 395
cagtatggac actgtccaaa gagaat 26

<210> 396
<211> 4090
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 396
aggaaauaag agagaaaaga agaguaagaa gaaauauaag agccaccaug agcgaggucg 60

aguucucuca cgaauauugg augagacacg cucucacccu ggcuaagaga gccagggacg 120

aaagagaggu gccaguuggc gcuguccugg uguugaacaa ucgcgucauc ggagaaggau 180

ggaaucgcgc cauuggccug cacgauccaa ccgcacaugc cgaaauuaug gcucugcggc 240

aaggcggccu cgugaugcaa aauuacagac uguacgaugc uacccucuac ggcaccuucg 300

agcccugugu caugugugcu ggggcaauga uucacucccg gauuggccgc gugguguuug 360

gagugcggaa ugccaagacu ggcgccgcug gaucucugau ggacguccug cacuauccug 420

ggaugaacca ccgggucgag aucacagagg gaauucuggc ugacgagugc gcagcccugc 480

ugugcgacuu cuauagaaug cccagaucgg uguucaacgc ccagaaaaaa gcucagagca 540

guaccaauuc cggcggaagc agcggaggau cuucuggaag cgaaacccca ggcaccagcg 600

agucugccac accagaauca ucuggcggua gcuccggcgg cuccaagaga aauuacaucc 660

ugggccucgc caucggcauc accucugucg gcuacggcau caucgacuac gagacacggg 720

augugauuga ugccggcgug cggcuguuca aagaggccaa cgucgagaac aacgagggcc 780

gcagaucuaa gaggggagcc agacggcuga agagaaggcg gagacacaga auccagagag 840

ugaagaagcu gcuguucgac uacaaccugc ugaccgacca cagcgagcug agcggcauca 900

acccuuauga ggccagagug aagggccuga gccagaagcu gagcgaggaa gaguuuagcg 960

ccgcacugcu gcaccuggcc aagagaagag gcgugcacaa cgugaacgaa guggaagagg 1020

acaccggcaa cgagcugucc accaaagagc agaucagcag aaacagcaag gcccuggaag 1080

agaaauacgu ggccgaacug cagcuggaac ggcugaaaaa ggauggcgaa gugcggggca 1140

gcaucaaccg guucaagacc agcgacuacg ugaaagaagc caaacagcug cugaaggugc 1200

agaaggccua ccaccagcug gaccagagcu ucaucgacac cuacaucgac cugcuggaaa 1260

cccggcggac cuacuaugaa ggaccuggcg agggaagccc cuucggcugg aaggacauca 1320

aagaauggua cgagaugcug augggccacu gcaccuacuu ucccgaggaa cugcggagcg 1380

ugaaguacgc cuacaacgcc gaccuguaca acgcccugaa cgaccugaac aaccucguga 1440

ucacccggga cgagaacgag aagcuggaau auuacgagaa guuccagauc aucgagaacg 1500

uguucaagca gaagaagaag cccacacuga agcagaucgc caaagagauc cuggucaacg 1560

aggaagauau caagggcuac agagugacca gcaccggcaa gcccgaguuc accaaccuga 1620

agguguacca cgacaucaag gauaucacag cccggaaaga gauuauugag aacgccgagc 1680

ugcuggacca aaucgccaag auccugacca ucuaccaguc cuccgaggac auccaagagg 1740

aacugaccaa ucugaacucc gagcugaccc aagaagagau cgagcagauc ucuaaucuga 1800

aggguacac aggcacccac aaccugagcc ugaaggccau caaccugauc cuggacgagc 1860

uguggcacac caacgacaac cagauugcca ucuucaaccg gcugaagcug gugcccaaga 1920

agguggaccu cagccagcag aaagaaaucc ccaccacacu gguggacgac uucauucuga 1980

gccccguggu caagagaagc uucauccaga gcaucaaagu gaucaacgcc aucaucaaga 2040

aguacgggcu gcccaacgau aucaucaucg agcuggcccg cgagaagaac uccaaggacg 2100

cucagaaaau gaucaacgag augcagaagc ggaaccggca gaccaacgag cggaucgagg 2160

aaaucauccg gaccaccggc aaagagaacg ccaaguaccu gaucgagaag aucaagcugc 2220

acgacaugca agagggcaag ugucuguaca gccuggaagc cauuccucug gaagaucugc 2280

ugaacaaucc cuucaacuac gagguggacc acaucauccc cagaagcgug uccuucgaca 2340

acagcuucaa caacaaggug cucgugaagc aagaggaaaa cuccaagaag ggcaacagaa 2400

ccccauucca guaccugagc agcuccgaca gcaagaucag cuacgaaacc uucaagaagc 2460

acauccugaa ucuggccaaa ggcaagggcc gcaucagcaa gaccaagaaa gaauaccugc 2520

ucgaggaacg ggacaucaac agauucagcg ugcagaaaga cuucaucaau cggaaccugg 2580

uggacaccag auacgccacc agaggccuga ugaaucugcu gagaagcuac uuccgcguga 2640

acaaucugga cgugaaaguc aaguccauca acggcggcuu caccagcuuu cugcggagaa 2700

aguggaaguu caagaaagag cggaacaagg gcuauaagca ccacgccgag gacgcccuga 2760

ucauugccaa cgccgauuuc aucuucaaag aguggaagaa acuggacaag gccaaaaaag 2820

ugauggaaaa ccagauguuc gaggaaaagc aggccgagag caugcccgag aucgaaaccg 2880

agcaagagua caaagagauu uucaucacgc cccaccagau caagcacauu aaggacuuca 2940

aggacuacaa guacagccac cgcguggaca agaagccuaa uagagagcug auuaacgaca 3000

cccuguacag cacccggaag gacgacaagg gcaauacccu gaucgucaac aaccugaacg 3060

gccuguacga caaggacaac gacaagcuca agaagcugau caacaagagc cccgagaaac 3120

ugcugaugua ccaccacgau ccucagaccu accagaaacu gaagcucauc auggaacagu 3180

acggcgacga gaagaauccc cuguacaagu acuacgagga aaccgggaac uaccugacca 3240

aguacuccaa aaaggacaau gggcccguga ucaagaagau uaaguauuac ggcaacaagc 3300

ugaaugccca ccuggacauc accgacgacu accccaacuc cagaaacaag guggucaagc 3360

ugucccugaa gccuuacaga uucgacgugu accuggacaa cggcguguac aaguucguga 3420

ccgugaagaa ccuggaugug aucaaaaaag aaaacuacua cgaagugaac agcaagugcu 3480

augaggaagc caagaaacuc aagaaaauca gcaaccaggc cgaguuuauc gccuccuucu 3540

acaacaacga ucugaucaag aucaacgggg agcuguauag agugauuggg gucaacaaug 3600

accugcugaa ccggaucgaa gucaacauga ucgacaucac cuaccgcgag uaccucgaga 3660

acaugaacga caagaggccu ccacggauca uuaagacaau cgccagcaag acgcagagca 3720

uuaagaagua cagcacugac auucugggca accuguacga agucaagagc aaaaagcacc 3780

cgcagauuau caagaaaggc gagggcgccg acaagagaac agccgauggu uccgaguucg 3840

aaagccccaa gaagaagagg aaagucuagu uaauuaagcu gccuucugcg gggcuugccu 3900

ucuggccaug cccuucuucu cucccuugca ccuguaccuc uuggucuuug aauaaagccu 3960

gaguaggaag aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4020

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4080

aaaaaaaaaa 4090

<210> 397
<211> 3822
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 397
atgagcgagg tcgagttctc tcacgaatat tggatgagac acgctctcac cctggctaag 60

agagccaggg acgaaagaga ggtgccagtt ggcgctgtcc tggtgttgaa caatcgcgtc 120

atcggagaag gatggaatcg cgccattggc ctgcacgatc caaccgcaca tgccgaaatt 180

atggctctgc ggcaaggcgg cctcgtgatg caaaattaca gactgtacga tgctaccctc 240

tacggcacct tcgagccctg tgtcatgtgt gctggggcaa tgattcactc ccggattggc 300

cgcgtggtgt ttggagtgcg gaatgccaag actggcgccg ctggatctct gatggacgtc 360

ctgcactatc ctgggatgaa ccaccgggtc gagatcacag agggaattct ggctgacgag 420

tgcgcagccc tgctgtgcga cttctataga atgcccagat cggtgttcaa cgcccagaaa 480

aaagctcaga gcagtaccaa ttccggcgga agcagcggag gatcttctgg aagcgaaacc 540

ccaggcacca gcgagtctgc cacaccagaa tcatctggcg gtagctccgg cggctccaag 600

agaaattaca tcctgggcct cgccatcggc atcacctctg tcggctacgg catcatcgac 660

tacgagacac gggatgtgat tgatgccggc gtgcggctgt tcaaagaggc caacgtcgag 720

aacaacgagg gccgcagatc taagagggga gccagacggc tgaagagaag gcggagacac 780

agaatccaga gagtgaagaa gctgctgttc gactacaacc tgctgaccga ccacagcgag 840

ctgagcggca tcaaccctta tgaggccaga gtgaagggcc tgagccagaa gctgagcgag 900

gaagagttta gcgccgcact gctgcacctg gccaagagaa gaggcgtgca caacgtgaac 960

gaagtggaag aggacaccgg caacgagctg tccaccaaag agcagatcag cagaaacagc 1020

aaggccctgg aagagaaata cgtggccgaa ctgcagctgg aacggctgaa aaaggatggc 1080

gaagtgcggg gcagcatcaa ccggttcaag accagcgact acgtgaaaga agccaaacag 1140

ctgctgaagg tgcagaaggc ctaccaccag ctggaccaga gcttcatcga cacctacatc 1200

gacctgctgg aaacccggcg gacctactat gaaggacctg gcgagggaag ccccttcggc 1260

tggaaggaca tcaaagaatg gtacgagatg ctgatgggcc actgcaccta ctttcccgag 1320

gaactgcgga gcgtgaagta cgcctacaac gccgacctgt acaacgccct gaacgacctg 1380

aacaacctcg tgatcacccg ggacgagaac gagaagctgg aatattacga gaagttccag 1440

atcatcgaga acgtgttcaa gcagaagaag aagcccacac tgaagcagat cgccaaagag 1500

atcctggtca acgaggaaga tatcaagggc tacagagtga ccagcaccgg caagcccgag 1560

ttcaccaacc tgaaggtgta ccacgacatc aaggatatca cagcccggaa agagattatt 1620

gagaacgccg agctgctgga ccaaatcgcc aagatcctga ccatctacca gtcctccgag 1680

gacatccaag aggaactgac caatctgaac tccgagctga cccaagaaga gatcgagcag 1740

atctctaatc tgaaggggta cacaggcacc cacaacctga gcctgaaggc catcaacctg 1800

atcctggacg agctgtggca caccaacgac aaccagattg ccatcttcaa ccggctgaag 1860

ctggtgccca agaaggtgga cctcagccag cagaaagaaa tccccaccac actggtggac 1920

gacttcattc tgagccccgt ggtcaagaga agcttcatcc agagcatcaa agtgatcaac 1980

gccatcatca agaagtacgg gctgcccaac gatatcatca tcgagctggc ccgcgagaag 2040

aactccaagg acgctcagaa aatgatcaac gagatgcaga agcggaaccg gcagaccaac 2100

gagcggatcg aggaaatcat ccggaccacc ggcaaagaga acgccaagta cctgatcgag 2160

aagatcaagc tgcacgacat gcaagagggc aagtgtctgt acagcctgga agccattcct 2220

ctggaagatc tgctgaacaa tcccttcaac tacgaggtgg accacatcat ccccagaagc 2280

gtgtccttcg acaacagctt caacaacaag gtgctcgtga agcaagagga aaactccaag 2340

aagggcaaca gaaccccatt ccagtacctg agcagctccg acagcaagat cagctacgaa 2400

accttcaaga agcacatcct gaatctggcc aaaggcaagg gccgcatcag caagaccaag 2460

aaagaatacc tgctcgagga acgggacatc aacagattca gcgtgcagaa agacttcatc 2520

aatcggaacc tggtggacac cagatacgcc accagaggcc tgatgaatct gctgagaagc 2580

tacttccgcg tgaacaatct ggacgtgaaa gtcaagtcca tcaacggcgg cttcaccagc 2640

tttctgcgga gaaagtggaa gttcaagaaa gagcggaaca agggctataa gcaccacgcc 2700

gaggacgccc tgatcattgc caacgccgat ttcatcttca aagagtggaa gaaactggac 2760

aaggccaaaa aagtgatgga aaaccagatg ttcgaggaaa agcaggccga gagcatgccc 2820

gagatcgaaa ccgagcaaga gtacaaagag attttcatca cgccccacca gatcaagcac 2880

attaaggact tcaaggacta caagtacagc caccgcgtgg acaagaagcc taatagagag 2940

ctgattaacg acaccctgta cagcacccgg aaggacgaca agggcaatac cctgatcgtc 3000

aacaacctga acggcctgta cgacaaggac aacgacaagc tcaagaagct gatcaacaag 3060

agccccgaga aactgctgat gtaccaccac gatcctcaga cctaccagaa actgaagctc 3120

atcatggaac agtacggcga cgagaagaat cccctgtaca agtactacga ggaaaccggg 3180

aactacctga ccaagtactc caaaaaggac aatgggcccg tgatcaagaa gattaagtat 3240

tacggcaaca agctgaatgc ccacctggac atcaccgacg actaccccaa ctccagaaac 3300

aaggtggtca agctgtccct gaagccttac agattcgacg tgtacctgga caacggcgtg 3360

tacaagttcg tgaccgtgaa gaacctggat gtgatcaaaa aagaaaacta ctacgaagtg 3420

aacagcaagt gctatgagga agccaagaaa ctcaagaaaa tcagcaacca ggccgagttt 3480

atcgcctcct tctacaacaa cgatctgatc aagatcaacg gggagctgta tagagtgatt 3540

ggggtcaaca atgacctgct gaaccggatc gaagtcaaca tgatcgacat cacctaccgc 3600

gagtacctcg agaacatgaa cgacaagagg cctccacgga tcattaagac aatcgccagc 3660

aagacgcaga gcattaagaa gtacagcact gacattctgg gcaacctgta cgaagtcaag 3720

agcaaaaagc acccgcagat tatcaagaaa ggcgagggcg ccgacaagag aacagccgat 3780

ggttccgagt tcgaaagccc caagaagaag aggaaagtct ag 3822

<210> 398
<211> 1273
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polypeptide sequence

<400> 398

```
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 Gly 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 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 Ser Val Phe Asn Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asn Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Lys Arg Asn Tyr Ile Leu Gly Leu Ala
        195                 200                 205

Ile Gly Ile Thr Ser Val Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg
```

```
    210                 215                 220

Asp Val Ile Asp Ala Gly Val Arg Leu Phe Lys Glu Ala Asn Val Glu
225                 230                 235                 240

Asn Asn Glu Gly Arg Arg Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg
                245                 250                 255

Arg Arg Arg His Arg Ile Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr
            260                 265                 270

Asn Leu Leu Thr Asp His Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu
        275                 280                 285

Ala Arg Val Lys Gly Leu Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser
    290                 295                 300

Ala Ala Leu Leu His Leu Ala Lys Arg Arg Gly Val His Asn Val Asn
305                 310                 315                 320

Glu Val Glu Glu Asp Thr Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile
                325                 330                 335

Ser Arg Asn Ser Lys Ala Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln
            340                 345                 350

Leu Glu Arg Leu Lys Lys Asp Gly Glu Val Arg Gly Ser Ile Asn Arg
        355                 360                 365

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

Gln Lys Ala Tyr His Gln Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile
385                 390                 395                 400

Asp Leu Leu Glu Thr Arg Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly
                405                 410                 415

Ser Pro Phe Gly Trp Lys Asp Ile Lys Glu Trp Tyr Glu Met Leu Met
            420                 425                 430
```

```
Gly His Cys Thr Tyr Phe Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala
        435                 440                 445

Tyr Asn Ala Asp Leu Tyr Asn Ala Leu Asn Asp Leu Asn Asn Leu Val
    450                 455                 460

Ile Thr Arg Asp Glu Asn Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln
465                 470                 475                 480

Ile Ile Glu Asn Val Phe Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln
                485                 490                 495

Ile Ala Lys Glu Ile Leu Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg
            500                 505                 510

Val Thr Ser Thr Gly Lys Pro Glu Phe Thr Asn Leu Lys Val Tyr His
        515                 520                 525

Asp Ile Lys Asp Ile Thr Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu
    530                 535                 540

Leu Leu Asp Gln Ile Ala Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu
545                 550                 555                 560

Asp Ile Gln Glu Glu Leu Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu
                565                 570                 575

Glu Ile Glu Gln Ile Ser Asn Leu Lys Gly Tyr Thr Gly Thr His Asn
            580                 585                 590

Leu Ser Leu Lys Ala Ile Asn Leu Ile Leu Asp Glu Leu Trp His Thr
        595                 600                 605

Asn Asp Asn Gln Ile Ala Ile Phe Asn Arg Leu Lys Leu Val Pro Lys
    610                 615                 620

Lys Val Asp Leu Ser Gln Gln Lys Glu Ile Pro Thr Thr Leu Val Asp
625                 630                 635                 640

Asp Phe Ile Leu Ser Pro Val Val Lys Arg Ser Phe Ile Gln Ser Ile
```

```
               645                 650                 655

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

Ile Ile Glu Leu Ala Arg Glu Lys Asn Ser Lys Asp Ala Gln Lys Met
        675                 680                 685

Ile Asn Glu Met Gln Lys Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu
    690                 695                 700

Glu Ile Ile Arg Thr Thr Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu
705                 710                 715                 720

Lys Ile Lys Leu His Asp Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu
                725                 730                 735

Glu Ala Ile Pro Leu Glu Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu
            740                 745                 750

Val Asp His Ile Ile Pro Arg Ser Val Ser Phe Asp Asn Ser Phe Asn
        755                 760                 765

Asn Lys Val Leu Val Lys Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg
    770                 775                 780

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

Thr Phe Lys Lys His Ile Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile
                805                 810                 815

Ser Lys Thr Lys Lys Glu Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg
            820                 825                 830

Phe Ser Val Gln Lys Asp Phe Ile Asn Arg Asn Leu Val Asp Thr Arg
        835                 840                 845

Tyr Ala Thr Arg Gly Leu Met Asn Leu Leu Arg Ser Tyr Phe Arg Val
    850                 855                 860
```

```
Asn Asn Leu Asp Val Lys Val Lys Ser Ile Asn Gly Gly Phe Thr Ser
865                 870                 875                 880


Phe Leu Arg Arg Lys Trp Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr
                885                 890                 895


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


Phe Lys Glu Trp Lys Lys Leu Asp Lys Ala Lys Lys Val Met Glu Asn
        915                 920                 925


Gln Met Phe Glu Glu Lys Gln Ala Glu Ser Met Pro Glu Ile Glu Thr
    930                 935                 940


Glu Gln Glu Tyr Lys Glu Ile Phe Ile Thr Pro His Gln Ile Lys His
945                 950                 955                 960


Ile Lys Asp Phe Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys
                965                 970                 975


Pro Asn Arg Glu Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp
            980                 985                 990


Asp Lys Gly Asn Thr Leu Ile Val  Asn Asn Leu Asn Gly  Leu Tyr Asp
        995                 1000                1005


Lys Asp  Asn Asp Lys Leu Lys  Lys Leu Ile Asn Lys  Ser Pro Glu
    1010                1015                1020


Lys Leu  Leu Met Tyr His His  Asp Pro Gln Thr Tyr  Gln Lys Leu
    1025                1030                1035


Lys Leu  Ile Met Glu Gln Tyr  Gly Asp Glu Lys Asn  Pro Leu Tyr
    1040                1045                1050


Lys Tyr  Tyr Glu Glu Thr Gly  Asn Tyr Leu Thr Lys  Tyr Ser Lys
    1055                1060                1065


Lys Asp  Asn Gly Pro Val Ile  Lys Lys Ile Lys Tyr  Tyr Gly Asn
```

```
    1070                1075                1080

Lys Leu  Asn Ala His Leu Asp  Ile Thr Asp Asp Tyr  Pro Asn Ser
    1085                1090                1095

Arg Asn  Lys Val Val Lys Leu  Ser Leu Lys Pro Tyr  Arg Phe Asp
    1100                1105                1110

Val Tyr  Leu Asp Asn Gly Val  Tyr Lys Phe Val Thr  Val Lys Asn
    1115                1120                1125

Leu Asp  Val Ile Lys Lys Glu  Asn Tyr Tyr Glu Val  Asn Ser Lys
    1130                1135                1140

Cys Tyr  Glu Glu Ala Lys Lys  Leu Lys Lys Ile Ser  Asn Gln Ala
    1145                1150                1155

Glu Phe  Ile Ala Ser Phe Tyr  Asn Asn Asp Leu Ile  Lys Ile Asn
    1160                1165                1170

Gly Glu  Leu Tyr Arg Val Ile  Gly Val Asn Asn Asp  Leu Leu Asn
    1175                1180                1185

Arg Ile  Glu Val Asn Met Ile  Asp Ile Thr Tyr Arg  Glu Tyr Leu
    1190                1195                1200

Glu Asn  Met Asn Asp Lys Arg  Pro Pro Arg Ile Ile  Lys Thr Ile
    1205                1210                1215

Ala Ser  Lys Thr Gln Ser Ile  Lys Lys Tyr Ser Thr  Asp Ile Leu
    1220                1225                1230

Gly Asn  Leu Tyr Glu Val Lys  Ser Lys Lys His Pro  Gln Ile Ile
    1235                1240                1245

Lys Lys  Gly Glu Gly Ala Asp  Lys Arg Thr Ala Asp  Gly Ser Glu
    1250                1255                1260

Phe Glu  Ser Pro Lys Lys Lys  Arg Lys Val
    1265                1270
```

<210> 399
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 399
attctctttg gacagtgtcc atactggtgg 30

<210> 400
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 400
taagagaaac ctgtcacagg tatgaccacc 30

<210> 401
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 401

Ile Leu Phe Gly Gln Cys Pro Tyr Trp Trp
1 5 10

<210> 402
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 402
ccaccagtat ggacactgtc caaagagaat 30

<210> 403
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic peptide sequence

<400> 403

Trp Trp Tyr Pro Cys Gln Gly Phe Leu Ile
1 5 10

<210> 404
<211> 108
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 404
caguauggac acuguccaaa guuuuaguac ucuguaauga aaauuacaga aucuacuaaa 60

acaaggcaaa augccguguu uaucucguca acuuguuggc gagauuuu 108

<210> 405
<211> 111
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 405
uuucaguaug gacacugucc aaaguuuuag uacucuguaa ugaaaauuac agaaucuacu 60

aaaacaaggc aaaaugccgu guuuaucucg ucaacuuguu ggcgagauuu u 111

<210> 406
<211> 100
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 406

caguauggac acuguccaaa guuuuaguac ucuggaaaca gaaucuacua aaacaaggca 60

aaaugccgug uuuaucucgu caacuuguug gcgagauuuu 100

<210> 407
<211> 109
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 407
ccaguaugga cacuguccaa aguuuuagua cucuguaaug aaaauuacag aaucuacuaa 60

aacaaggcaa aaugccgugu uuaucucguc aacuuguugg cgagauuuu 109

<210> 408
<211> 110
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 408
accaguaugg acacugucca aaguuuuagu acucuguaau gaaaauuaca gaaucuacua 60

aaacaaggca aaaugccgug uuuaucucgu caacuuguug gcgagauuuu 110

<210> 409
<211> 111
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 409
caccaguaug gacacugucc aaaguuuuag uacucuguaa ugaaaauuac agaaucuacu 60

aaaacaaggc aaaaugccgu guuuaucucg ucaacuuguu ggcgagauuu u 111

<210> 410
<211> 112
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 410
ccaccaguau ggacacuguc caaaguuuua guacucugua augaaaauua cagaaucuac 60

uaaaacaagg caaaaugccg uguuuaucuc gucaacuugu uggcgagauu uu 112

<210> 411
<211> 104
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 411
caguauggac acuguccaaa guuuuaguac ucuguagaaa uacagaaucu acuaaaacaa 60

ggcaaaaugc cguguuuauc ucgucaacuu guuggcgaga uuuu 104

<210> 412
<211> 104
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 412
caguauggac acuguccaaa guuuuaguac ucugcggaaa cgcagaaucu acuaaaacaa 60

ggcaaaaugc cguguuuauc ucgucaacuu guuggcgaga uuuu 104

<210> 413
<211> 96
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 413
caguauggac acuguccaaa guuuuaguac ucgaaagaau cuacuaaaac aaggcaaaau 60

gccguguuua ucucgucaac uuguuggcga gauuuu 96

<210> 414
<211> 96
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 414
caguauggac acuguccaaa guuuuaguac ccgaaagcau cuacuaaaac aaggcaaaau 60

gccguguuua ucucgucaac uuguuggcga gauuuu 96

<210> 415
<211> 108
<212> RNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic polynucleotide sequence

<400> 415
caguauggac acuguccaaa guuuuaguac ucuguaauga aaauuacaga aucuacuaaa 60

acaaggcaaa augccguguu uaucucguca acuuguuggc gagauuuu 108

<210> 416
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 416
ccaccagtat ggacactgtc caaa 24

<210> 417
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 417
caccagtatg gacactgtcc aaag 24

<210> 418
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 418
accagtatgg acactgtcca aaga 24

<210> 419
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 419
ccagtatgga cactgtccaa agag 24

<210> 420
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 420
cagtatggac actgtccaaa gaga 24

<210> 421
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 421
agtatggaca ctgtccaaag aga 23

<210> 422
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 422
gtatggacac tgtccaaaga gaat 24

<210> 423
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 423
tatggacact gtccaaagag aatc 24

<210> 424
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 424
acctactgat gatgcacctt tgatcaatag at 32

<210> 425
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 425
catcacccct cgggatggtt ctt 23

<210> 426
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 426
cagcccagaa tcccaaccac aaaat 25

<210> 427
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of artificial sequence: synthetic oligonucleotide sequence

<400> 427
agaccagctc gacttgggat gg 22