



US 20150165054A1

(19) **United States**(12) **Patent Application Publication**
Liu et al.(10) **Pub. No.: US 2015/0165054 A1**(43) **Pub. Date: Jun. 18, 2015**(54) **METHODS FOR CORRECTING CASPASE-9
POINT MUTATIONS**(71) Applicant: **President and Fellows of Harvard
College, Cambridge, MA (US)**(72) Inventors: **David R. Liu, Lexington, MA (US);
Alexis Christine Komor, Pasadena, CA
(US)**(73) Assignee: **President and Fellows of Harvard
College, Cambridge, MA (US)**(21) Appl. No.: **14/326,303**(22) Filed: **Jul. 8, 2014****Related U.S. Application Data**(60) Provisional application No. 61/980,333, filed on Apr.
16, 2014, provisional application No. 61/915,386,
filed on Dec. 12, 2013.**Publication Classification**(51) **Int. Cl.**
A61K 47/48 (2006.01)
A61K 38/50 (2006.01)
A61K 38/46 (2006.01)
C12P 19/34 (2006.01)(52) **U.S. Cl.**CPC **A61K 47/4823** (2013.01); **C12P 19/34**
(2013.01); **A61K 38/50** (2013.01); **A61K**
38/465 (2013.01); **C12Y 305/04004** (2013.01);
C12Y 305/04005 (2013.01); **C12Y 301/22**
(2013.01)(57) **ABSTRACT**

Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for the targeted editing of nucleic acids, including editing a nucleic acid encoding a mutant Caspase-9 protein to correct a point mutation associated with a disease or disorder, e.g., with neuroblastoma. The methods provided are useful for correcting a Caspase-9 point mutation within the genome of a cell or subject, e.g., within the human genome. In some embodiments, fusion proteins of Cas9 and nucleic acid editing enzymes or enzyme domains, e.g., deaminase domains, are provided. In some embodiments, reagents and kits for the generation of targeted nucleic acid editing proteins, e.g., fusion proteins of Cas9 and nucleic acid editing enzymes or domains, are provided.

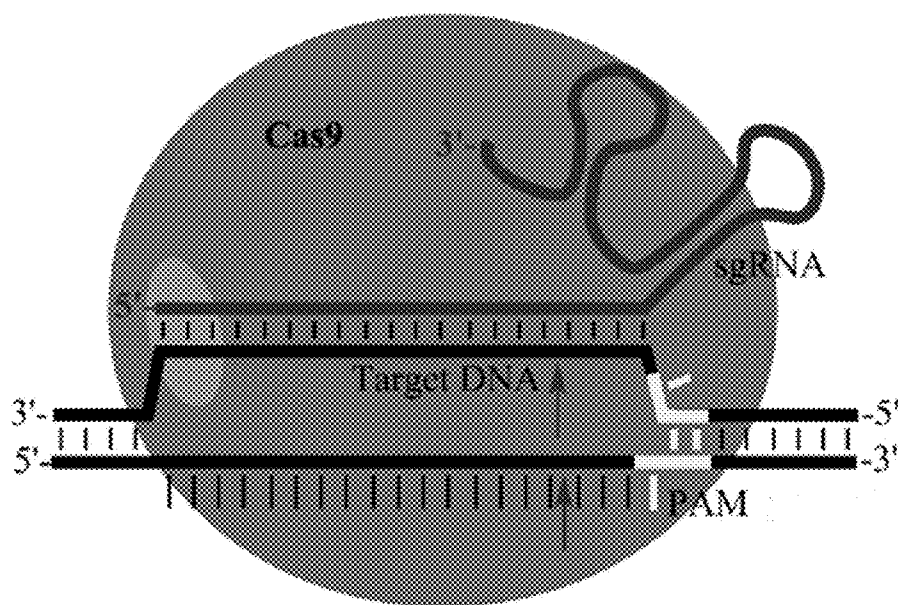


FIGURE 1

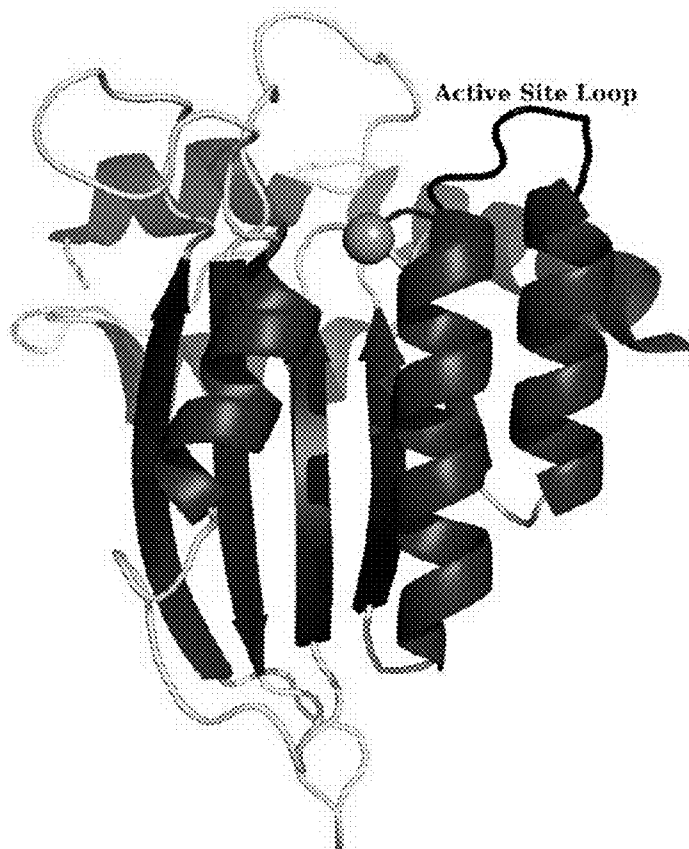


FIGURE 2

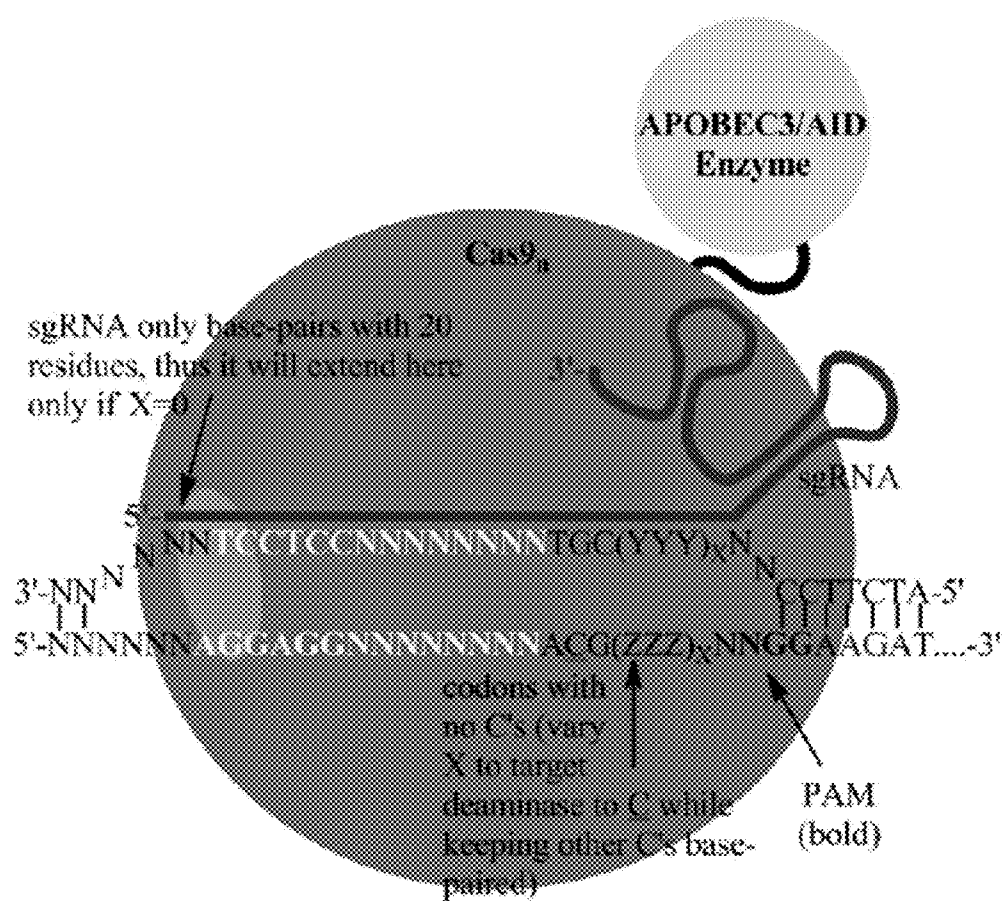


FIGURE 3

METHODS FOR CORRECTING CASPASE-9 POINT MUTATIONS

RELATED APPLICATION

[0001] This application claims priority under 35 U.S.C. §119(e) to U.S. provisional patent application, U.S. Ser. No. 61/915,386 filed Dec. 12, 2013, and U.S. provisional patent application, U.S. Ser. No. 61/980,333 filed Apr. 16, 2014, the entire contents of each of which is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Targeted editing of nucleic acid sequences, for example, the introduction of a specific modification into genomic DNA, is a highly promising approach for the study of gene function and also has the potential to provide new therapies for human genetic diseases.¹ An ideal nucleic acid editing technology possesses three characteristics: (1) high efficiency of installing the desired modification; (2) minimal off-target activity; and (3) the ability to be programmed to edit precisely any site in a given nucleic acid, e.g., any site within the human genome.² Current genome engineering tools, including engineered zinc finger nucleases (ZFNs),³ transcription activator like effector nucleases (TALENs),⁴ and most recently, the RNA-guided DNA endonuclease Cas9,⁵ effect sequence-specific DNA cleavage in a genome. This programmable cleavage can result in mutation of the DNA at the cleavage site via non-homologous end joining (NHEJ) or replacement of the DNA surrounding the cleavage site via homology-directed repair (HDR).^{6,7}

[0003] One drawback to the current technologies is that both NHEJ and HDR are stochastic processes that typically result in modest gene editing efficiencies as well as unwanted gene alterations that can compete with the desired alteration.⁸ Since many genetic diseases in principle can be treated by effecting a specific nucleotide change at a specific location in the genome (for example, a C to T change in a specific codon of a gene associated with a disease),⁹ the development of a programmable way to achieve such precision gene editing would represent both a powerful new research tool, as well as a potential new approach to gene editing-based human therapeutics.

SUMMARY OF THE INVENTION

[0004] The clustered regularly interspaced short palindromic repeat (CRISPR) system is a recently discovered prokaryotic adaptive immune system¹⁰ that has been modified to enable robust and general genome engineering in a variety of organisms and cell lines.¹¹ CRISPR-Cas (CRISPR associated) systems are protein-RNA complexes that use an RNA molecule (sgRNA) as a guide to localize the complex to a target DNA sequence via base-pairing.¹² In the natural systems, a Cas protein then acts as an endonuclease to cleave the targeted DNA sequence.¹³ The target DNA sequence must be both complementary to the sgRNA, and also contain a “protospacer-adjacent motif” (PAM) dinucleotide at the 3'-end of the complementary region in order for the system to function (FIG. 1).¹⁴ Among the known Cas proteins, *S. pyogenes* Cas9 has been mostly widely used as a tool for genome engineering.¹⁵ This Cas9 protein is a large, multi-domain protein containing two distinct nuclease domains. Point mutations can be introduced into Cas9 to abolish nuclease activity, resulting in a dead Cas9 (dCas9) that still retains its

ability to bind DNA in a sgRNA-programmed manner.¹⁶ In principle, when fused to another protein or domain, dCas9 can target that protein to virtually any DNA sequence simply by co-expression with an appropriate sgRNA.

[0005] The potential of the dCas9 complex for genome engineering purposes is immense. Its unique ability to bring proteins to specific sites in a genome programmed by the sgRNA in theory can be developed into a variety of site-specific genome engineering tools beyond nucleases, including transcriptional activators, transcriptional repressors, histone-modifying proteins, integrases, and recombinases.¹¹ Some of these potential applications have recently been implemented through dCas9 fusions with transcriptional activators to afford RNA-guided transcriptional activators,^{17,18} transcriptional repressors,^{16,19,20} and chromatin modification enzymes.²¹ Simple co-expression of these fusions with a variety of sgRNAs results in specific expression of the target genes. These seminal studies have paved the way for the design and construction of readily programmable sequence-specific effectors for the precise manipulation of genomes.

[0006] Significantly, 80-90% of protein mutations responsible for human disease arise from the substitution, deletion, or insertion of only a single nucleotide.⁶ No genome engineering tools, however, have yet been developed that enable the manipulation of a single nucleotide in a general and direct manner. Current strategies for single-base gene correction include engineered nucleases (which rely on the creation of double-strand breaks, DSBs, followed by stochastic, inefficient homology-directed repair, HDR), and DNA-RNA chimeric oligonucleotides.²² The latter strategy involves the design of a RNA/DNA sequence to base pair with a specific sequence in genomic DNA except at the nucleotide to be edited. The resulting mismatch is recognized by the cell's endogenous repair system and fixed, leading to a change in the sequence of either the chimera or the genome. Both of these strategies suffer from low gene editing efficiencies and unwanted gene alterations, as they are subject to both the stochasticity of HDR and the competition between HDR and non-homologous end-joining, NHEJ.²³⁻²⁵ HDR efficiencies vary according to the location of the target gene within the genome,²⁶ the state of the cell cycle,²⁷ and the type of cell/tissue.²⁸ The development of a direct, programmable way to install a specific type of base modification at a precise location in genomic DNA with enzyme-like efficiency and no stochasticity would therefore represent a powerful new approach to gene editing-based research tools and human therapeutics.

[0007] Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for the targeted editing of nucleic acids, including editing a single site within a subject's genome, e.g., the human genome. In some embodiments, fusion proteins of Cas9 and nucleic acid editing enzymes or enzyme domains, e.g., deaminase domains, are provided. In some embodiments, methods for targeted nucleic acid editing are provided. In some embodiments, reagents and kits for the generation of targeted nucleic acid editing proteins, e.g., fusion proteins of Cas9 and nucleic acid editing enzymes or domains, are provided.

[0008] Some aspects of this disclosure provide fusion proteins comprising (i) a nuclease-inactive CAS9 domain; and (ii) a nucleic acid-editing domain. In some embodiments, the nucleic acid-editing domain is a DNA-editing domain. In some embodiments, the nucleic-acid-editing domain is a deaminase domain. In some embodiments, the deaminase is a

cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC1 family deaminase. In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). In some embodiments, the deaminase is an ACF1/ASE deaminase. In some embodiments, the deaminase is an adenosine deaminase. In some embodiments, the deaminase is an ADAT family deaminase. In some embodiments, the nucleic-acid-editing domain is fused to the N-terminus of the CAS9 domain. In some embodiments, the nucleic-acid-editing domain is fused to the C-terminus of the CAS9 domain. In some embodiments, the CAS9 domain and the nucleic-acid-editing domain are fused via a linker. In some embodiments, the linker comprises a (GGGS)_n (SEQ ID NO: 91), a (G)_n, an (EAAAK)_n (SEQ ID NO: 5), or an (XP)_n motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

[0009] Some aspects of this disclosure provide methods for DNA editing. In some embodiments, the methods comprise contacting a DNA molecule with (a) a fusion protein comprising a nuclease-inactive Cas9 domain and a deaminase domain; and (b) an sgRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the sgRNA in an amount effective and under conditions suitable for the deamination of a nucleotide base. In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleotide base results in a sequence that is not associated with a disease or disorder. In some embodiments, the DNA sequence comprises a T>C or A>G point mutation associated with a disease or disorder, and wherein the deamination of the mutant C or G base results in a sequence that is not associated with a disease or disorder. In some embodiments, the deamination corrects a point mutation in the sequence associated with the disease or disorder. In some embodiments, the sequence associated with the disease or disorder encodes a protein, and wherein the deamination introduces a stop codon into the sequence associated with the disease or disorder, resulting in a truncation of the encoded protein. In some embodiments, the deamination corrects a point mutation in the PI3KCA gene, thus correcting an H1047R and/or a A3140G mutation. In some embodiments, the contacting is performed in vivo in a subject susceptible to having, having, or diagnosed with the disease or disorder. In some embodiments, the disease or disorder is a disease associated with a point mutation, or a single-base mutation, in the genome. In some embodiments, the disease is a genetic disease, a cancer, a metabolic disease, or a lysosomal storage disease.

[0010] Some aspects of this disclosure provide a reporter construct for detecting nucleic-acid-editing activity of a Cas9:DNA-editing domain fusion protein. In some embodiments, the construct comprises (a) a reporter gene comprising a target site for the Cas9 DNA-editing protein, wherein targeted DNA editing results in an increase in expression of the reporter gene; and (b) a promoter sequence that controls expression of the reporter gene. In some embodiments, the construct further comprises (c) a sequence encoding an sgRNA targeting the Cas9 DNA-editing protein to the target site of the reporter gene, wherein expression of the sgRNA is independent of the expression of the reporter gene. In some embodiments, the target site of the reporter gene comprises a

premature stop codon, and wherein targeted DNA editing of the template strand by the Cas9 DNA-editing protein results in a conversion of the premature stop codon to a codon encoding an amino acid residue. In some embodiments, the reporter gene encodes a luciferase, a fluorescent protein, or an antibiotic resistance marker.

[0011] Some aspects of this disclosure provide kits comprising a nucleic acid construct that comprises a sequence encoding a nuclease-inactive Cas9 sequence, a sequence comprising a cloning site positioned to allow cloning of a sequence encoding a nucleic acid-editing enzyme or enzyme domain in-frame with the Cas9-encoding sequence, and, optionally, a sequence encoding a linker positioned between the Cas9 encoding sequence and the cloning site. In addition, in some embodiments, the kit comprises suitable reagents, buffers, and/or instructions for in-frame cloning of a sequence encoding a nucleic acid-editing enzyme or enzyme domain into the nucleic acid construct to generate a Cas9 nucleic acid editing fusion protein. In some embodiments, the sequence comprising the cloning site is N-terminal of the Cas9 sequence. In some embodiments, the sequence comprising the cloning site is C-terminal of the Cas9 sequence. In some embodiments, the encoded linker comprises a (GGGS)_n (SEQ ID NO: 91), a (G)_n, an (EAAAK)_n (SEQ ID NO: 5), or an (XP)_n motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

[0012] Some aspects of this disclosure provide kits comprising a fusion protein comprising a nuclease-inactive Cas9 domain and a nucleic acid-editing enzyme or enzyme domain, and, optionally, a linker positioned between the Cas9 domain and the nucleic acid-editing enzyme or enzyme domain. In addition, in some embodiments, the kit comprises suitable reagents, buffers, and/or instructions for using the fusion protein, e.g., for in vitro or in vivo DNA or RNA editing. In some embodiments, the kit comprises instructions regarding the design and use of suitable sgRNAs for targeted editing of a nucleic acid sequence.

[0013] The summary above is meant to illustrate, in a non-limiting manner, some of the embodiments, advantages, features, and uses of the technology disclosed herein. Other embodiments, advantages, features, and uses of the technology disclosed herein will be apparent from the Detailed Description, the Drawings, the Examples, and the Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] FIG. 1. The Cas9/sgRNA-DNA complex. The 3' end of the sgRNA forms a ribonucleoprotein complex with the Cas9 nuclease, while the 20 nt 5' end of the sgRNA recognizes its complementary stretch of DNA. DNA binding requires the 3-nt PAM sequence 5' to the target DNA. In the case of wtCas9, double-strand DNA cleavage occurs 3 nt from the PAM to produce blunt ends (shown by the arrows). It should be noted that the size of the bubble is unknown.

[0015] FIG. 2. Crystal structure of the catalytic domain of APOBEC3G (PDB ID 3E1U). The core secondary structure, which is believed to be conserved among the entire family, consists of a five-stranded β -sheet (arrows) flanked by six α -helices. The active center loop (active site loop), is believed to be responsible for determining deamination specificity. The Zn²⁺ responsible for catalytic activity is shown as a sphere.

[0016] FIG. 3. Design of luciferase-based reporter assay. The sgRNA will be varied to target numerous sequences that correspond to regions prior to and including the luciferase

gene in order to target the mutated start codon (C residue underlined). A “buffer” region will be added between the start codon and the luciferase gene to include codons of only A’s and T’s (shown as (ZZZ)_x). The Shine-Dalgarno sequence is indicated. In some embodiments, it is preferable to keep all C’s base-paired to prevent off-target effects.

DEFINITIONS

[0017] As used herein and in the claims, the singular forms “a,” “an,” and “the” include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to “an agent” includes a single agent and a plurality of such agents.

[0018] The term “Cas9” or “Cas9 nuclease” refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (e.g., a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 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., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *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., J. J., McShan W. M., Ajdic D. J., Savic D. J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A. N., Kenton S., Lai H. S., Lin S. P., Qian Y., Jia H. G., Najjar F. Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S. W., Roe B. A., McLaughlin R. E., 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., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., Eckert M. R., Vogel J., Charpentier E., *Nature* 471:602-607 (2011); and “A programmable dual-RNA-guided DNA endo-

nuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *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, a Cas9 nuclease has an inactive (e.g., an inactivated) DNA cleavage domain.

[0019] A nuclease-inactivated Cas9 protein may interchangeably be referred to as a “dCas9” protein (for nuclease-“dead” 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 H841A 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, 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% 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% to the corresponding fragment of wild type Cas9. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, SEQ ID NO:1 (nucleotide); SEQ ID NO:2 (amino acid)).

(SEQ ID NO: 1)

```
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAA
GGTTCGGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTT
TATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAACGGACAGCTCGTAGAAGGTATACACCGTCCGAAGAAT
CGTATTTGTTATCTACAGGAGATTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGA
```

-continued

GTCTTTTTTGGTGAAGAAGACAGAAGCATGAACGTCATCCTATTTTGGAAATATAGTAGAGTAAGTTGCTTATC
 ATGAGAAATATCCAACCTATCATCTGCGAAAAAATTGGCAGATTCTACTGATAAAGCGGATTGCGCTTAATC
 TATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTAAATCCTGATAATAGTGA
 TGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATATTTTGAAGAAAAACCTATTAAACGCAAGTAGAG
 TAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAAATCTATTGCTCAGCTCCCGGT
 GAGAAGAGAAATGGCTTGTGGGAATCTCATTGCTTGTGATTGGGATTGACCCCTAATTTAAATCAAAATTTGA
 TTTGGCAGAAGATGCTAAATACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATGGCGCAAAATG
 GAGATCAATATGCTGATTTGTTTTGGCAGCTAAGAATTTATCAGATGCTATTTACTTTTTCAGATATCCTAAGAGTA
 AATAGTGAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCT
 TTTAAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTTTTGATCAATCAAAAAACGATATG
 CAGGTATATATGATGGGGAGCTAGCCAAGAGAATTTATAAATTTATCAAAACCAATTTTAGAAAAAATGGATGGT
 ACTGAGGAATTTATTTGGTGAACATAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCC
 CCATCAAAATTCATTGGGTGAGCTGCTGCTATTTTGAGAAGACAAGAGACTTTATCCATTTTAAAAGACAATC
 GTGAGAAGATTGAAAAATCTTGACTTTTCCGAATTCCTTATATGTTGGTCCATTGGCGCTGGCAATAGTCGTTTT
 GCATGGATGACTCGGAAGCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGC
 TCAATCATTTATTGAACGCATGACAACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGC
 TTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAATATGTTACTGAGGGAATGCGAAAACAGCATTT
 CTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTACTCTTCAAAACAAATCGAAAAAGTAACCGTTAAGCAATTA
 AGAAGATTATTTCAAAAAATAGAAATGTTTTGATAGTGTGAAATTTTCAAGGATTGAAGATAGATTTAATGCTTCAT
 TAGGCGCTTACCATGATTTGCTAAAAATTTATAAAGATAAAGATTTTTTGGATAATGAAGAAAAAGAGATATCTTA
 GAGGATATTTTAAACATTGACCTTATTTGAAGATAGGGGATGATTGAGGAAAGACTTAAAAACATATGCTCACCT
 CTTTGATGATAAGGTGATGAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAAATGATTA
 ATGGTATTAGGGATAAGCAATCTGGCAAAACAATATAGATTTTTTGAAATCAGATGGTTTGGCAATCGCAATTTT
 ATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAGATATTTCAAAAGCACAGGTGCTGCGACAGGCCATAG
 TTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCGCTATTTAAAAAGGTATTTTACAGACTGTAAAAATTTGTTG
 ATGAACCTGGTCAAGTAAATGGGGCATAAGCCAGAAAAATATCGTTATTGAAATGGCAGCTGAAAAATCAGACAACCTAA
 AAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAGAATTAGGAAGTCAGATCTTAA
 AGAGCATCTGTTGAAAAATCTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAAATGGAAGAGACATGT
 ATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTATTATGATGTCGATCACATTGTTCCACAAGTTTCATTAAA
 GAGGATCTAATAGACAAATAAGGTACTAACCGCTTCGATAAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGA
 AGTAGTCAAAAGATGAAAACTATTGGAGACAACCTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATT
 TAACGAAAGCTGAACGTGGAGGTTTGAAGTAACTTGATAAAGCTGGTTTTATCAAAACGCCAATTGGTTGAACTCGC
 CAAATCACTAAGCATGTGGCAGCAAAATTTGGATAGTCGCATGAATACTAAATACGATGAAATGATAAACTATTTCG
 AGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTCCAATCTATAAAGTACGTG
 AGATTAACAATTACCATTCGCCCATGATGCGTATCTAAATGCCGCTCGTTGGAACGCTTTGATTAAAGAAATATCCA
 AACTTGAATCGGAGTTTGTCTTGGTGATTATAAAGTTTATGATGTTCTGTAATGATTGCTAAGTCTGAGCAAGA
 AATAGGCAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACCTCTTCAAAACAGAAATTACACTTGCAA
 ATGGAGAGATTCCGCAACGCCCTCTAATCGAACTAATGGGGAACCTGGAGAAATGTCTGGGATAAAGGGCGAGAT
 TTTGCCACAGTGCAGAAAGTATTGTCATGCCCAAGTCAATATTGTCAAGAAAAAGAGTACAGACAGGCGGAT
 CTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGAACAAGCTTATGCTCGTAAAAAAGACTGGGATCCAAAAAAT
 ATGGTGGTTTTGATAGTCCAAAGGTAGCTTATTCAGTCTAGTGGTTGCTAAGGTGAAAAAAGGAAATCGAAGAAG
 TTAATAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTT
 AGAAGCTAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAA
 ACGGTCGTAACCGGATCTGGCTAGTCCGGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAGCAAAATATGTG
 AATTTTTTATATTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTGTG
 GGAGCAGCATAGCATATTATTAGATGAGATTATTGAGCAAAATCAGTGAATTTTCTAAGCGTGTATTTTAGCAGATG
 CCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAAATATTATT
 CTTATTATTACGTTGACGAATCTTGAGCTCCCGCTGCTTTAAATATTTTGATACAACAATTGATCGTAAACGATA
 TACGCTACAAAAAGAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGCTTTTATGAAACACGCATTGATT
 TGAGTCAGCTAGGAGGTGACTGA

(SEQ ID NO: 2)

MDKKYSIGLDIGTNSVGVAVITDDYKVPKSKFKVLGNTDRHSIKKNLIGALLFGSGETAETRLKRTARRRYTRRN
 RICYLQEI FSNEMAKVDDSPFHRLEESFLVEEDKKHERHP IFGNI VDEVAYHEKYPTI YHLRKKLADSTDKADLRLI
 YLALAHMI KPRGHFLIEGLNPDNSVDKLFILQVQIYNQLFEENPINASRVDAKAIL SARLSKSRLENLIAQLPG
 EKRNLGFLGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLF LAAKNLSDAI LLSDLIRV
 NSEITKAPLSASMI KRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEFYKFKIPILEKMDG
 TEELLVLKLNREDLLRQRTFDNGSIPHQIHLGELHALRRQEDFPYFLKDNREKIEKILTRFIPYVVGPLRGNRSF
 AWMTKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLEYFTVYNELTKVKYVTEGMRKPAF
 LSGEQKAIIVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRFNALSGLAYHDLKIIKD KDFLDNEENEDIL
 EDIVLTLLTFEDRGMIEERLKYAHLFPDDKVMKQLKRRRYTGWGRLSRKLINGIRDQSGKTI LDFLKS DGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQHSLSHEQIANLAGSPA IKKGILQTVKIVDELVKVMGHKPENIV IEMARENQTQ
 GKQKNSRERMKRIEIEGIELKELSGQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYVDHIVPQSF I K
 DDISDKVLTRSDKNRKSNDPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
 QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDPRKDFQFYKREINNYHHAHDAYLNAVVGTA LIKKYP
 KLESEFVYGDYKIVYDVVKMI AKSBOEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRLPIETNGETGEIVWDKGRD
 FATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFPDPTVAYSVLVVAKEGKSKK
 LKSVKELLGTTIMERSEFELNPDIFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
 NFLYLASHYEKLGKSPEDNEQQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENII
 HLFPTLNLGAPAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD
 (single underline: HNH domain; double underline: RuvC domain)

[0020] In some embodiments, wild type Cas9 corresponds to, or comprises SEQ ID NO: 3 (nucleotide) and/or SEQ ID NO: 4 (amino acid):

(SEQ ID NO: 3)

ATGGATAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCGGTTGGATGGGCTGTCATAACCGATGAATACAA
AGTACCTTCAAAGAAATTTAAGGTGTTGGGGAAACACAGACCGTCATTTCGATTAAAAAGAAATCTTATCGGTGCCCTCC
TATTTCGATAGTGGCGAAACGCGACGAGCGACTCGCCTGAAACGAACCGCTCGGAGAAAGGTATACACGTGCGCAAGAAC
CGAATATGTTACTTACAAGAAATTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTCCACCGTTTGGGAAGA
GTCTTCTTGTGCGAAGAGGACAGAAACATGAACGGCACCCCATCTTTGGAACATAGTAGATGAGGTGGCATATC
ATGAAAAGTACCCACCGATTATCACCTCAGAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATC
TACTTGGCTCTTGCCCATATGATAAAGTTCGGTGGGCATTCTTCATTGAGGGTGATCTAAATCCGGACAACCTCGGA
TGTCGACAAACTGTTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCTATAAATGCAAGTGGCG
TGATGCGAAAGGTATTCTTTAGCCGCCCTCTCTAAATCCCGACGGCTAGAAACCTGATCGCACAAATACCCGGA
GAGAAGAAAAATGGGTTGTTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAAATTTAAGTCGAACCTCGA
CTTAGCTGAAGATGCCAAATTCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCTACTGGCACAAATG
GAGATCAGTATGCGGACTTATTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCTCTATCTGACATACTGAGATT
AATACTGAGATTACCAAGGCGCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACT
TCTCAAGGCCCTAGTCGTCAGCAACTGCTGAGAAATATAAGGAAATATTCTTGTATCAGTCGAAAAACGGGTACG
CAGGTTATATTGACGCGGGGCGAGTCAAGAGGAATTCACAAGTTTATCAAACCATATTAGAGAAGATGATGGG
ACGGAAGAGTTGCTTGTAAGAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCC
ACATCAAATCCATTAGGCGAATTCATGCTATACCTTAGAAGGACAGGAGGATTTTATCCGTTCTCAAAGACAATC
TGAAAGAGATTGAGAAAAATCCGAACTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACCTCGGTT
GCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTGAGGAAGTTGTCGATAAAGGTGCGTCAGC
TCAATCGTTTCATCGAGAGGATGACCAACTTTGACAAGAAATTCACGAACGAAAAAGTATTGCTTAAGCACAGTTTAC
TTTACGAGTATTTCCACAGTGTACAACTGAACTCAGAAAGTTAAGTATGTCACTGAGGGCATCGTAAACCCGCTTT
CTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAGTGACAGTTAAGCAATTGAA
AGAGGACTACTTTAAGAAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGTTAGAAGATCGATTAAATGCGTCAC
TTGGTACGTATCATGACCTCTTAAAGATAATTAAAGATAAGGACTTCTCGATAACGAAGAAATGAAGATATCTTA
GAAGATATAGTGTGACTCTTACCTCTTTGAAGATCGGGAATGATTGAGGAAGACTAAAAACATACGCTCACCT
GTTTCGACGATAAGGTTATGAACAGTTTAAAGAGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCA
ACGGGATAAGAGACAAAGTAAAGTGTGTAACCTATTCGATTTCCTAAGAGCGACGGCTTCGCAATAGGAACCTT
ATGCAGCTGATCCATGATGACTCTTAACTTCAAAGAGGATATACAAAGGCACAGGTTTCCGGACAAGGGGACTC
ATTGCGACGAACATATTGCGAATCTTGTGCTTCCGACGCCATCAAAAGGGCATACTCCAGACAGTCAAAGTAGTGG
ATGAGCTAGTTAAGGTTCAGGACTGGACATCAACCGGAAACATTGTAATCGAGATGGCAGCGGAAAAATCAACGACT
CAGAAGGGGCAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCGCCAGATCTT
AAAGGAGCATCTGTGGAATAATCCCAATTGCGAAGACGAGAACTTTACCTCTATTACCTACAAATGGAAGGGACA
TGATGTTGATCAGGAATCGACATGACATCAACCGTTTATCTGATTACGACGTCGATCACATTGTACCCCAATCTTTT
AAGGACGATTCAATCGACAATAAGTGCTTACACGCTCGGATAAGAACCGAGGAAAAAGTGACAAATGTTCCAAGCGA
GGAAGTCGTAAGAAAAATGAAGAACTATTGGCGGCAGCTCTTAATGCGAACTGATAACGCAAGAAAGTTTCGATA
ACTTAACATAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTATTAAACGTCAGCTCGTGGAAACC
CGCCAAATCACAAAGCATGTTGACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGAT
TCGGGAAGTCAAAGTAACTACTTTAAAGTCAAAATGGTGTGCGACTTCAGAAAGGATTTTCAATTCTATAAAGTTA
GGGAGATAAATAACTACCACTACCGCGCACGACGCTTATCTTAATGCCGTGCTAGGACCGCACTCATTAAGAAATAC
CCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCGTAAGATGATCGCGAAAGCGAACA
GGAGATAGGCAAGGCTACAGCCAAATCTCTTTATTTCTAACATTATGAATTTCTTTAAGACGGAATCACTCTGG
CAACCGGAGAGATACGCAAAACGCTTTAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGG
GACTTCGCGACGGTGAGAAAAAGTTTGTCCATGCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGG
GTTTTCAAAGGAATCGATTCTTCAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCGGAAAA
AGTAGGTGGCTTCGATAGCTCAGTTGCTTATCTGCTTAGTAGTGGCAAAAGTTGAGAAGGAAAAATCCAG
AACTGAAGTCAGTCAAGAAATTTATGGGATAACGATTATGGAGCGCTCGTCTTTTGAAGAAGACCCCATCGACTT
CCTTGAGGCGAAAGGTTACAAAGGAAGTAAAAAGGATCTCATAATTAACTACCAAGTATAGTCTGTTTGAGTTAG
AAAATGGCGGAAACAGATGTTGGCTAGCGCCGAGAGCTTCAAAGGGGAACGAACTCGCACTACCGCTCTAAATAC
GTGAATTTCTGTATTAGCGTCCCATACGAGAAGTTGAAAGGTTACCTGAAGATAACGAACAGAAGCAACTTTT
TGTTGAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAATTTCGGAATTCAGTAAGAGAGTCATCTAGCTG
ATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGATAAACCCATACGTGAGCAGGCGGAAAAATAT
ATCCATTGTTTACTCTTACCAACCTCGCGCTCCAGCGCATTCAGTATTTTGACACAACGATAGATCGCAACG
ATACACTTCTACCAAGAGGTGCTAGACGCGACACTGATTACCAATCCATCACGGGATTATATGAACTCGGATAG
ATTTGTCACAGCTTGGGGTGACGGATCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAGACCATGACGGT
GATTATAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA

(SEQ ID NO: 4)

MDKKYISGLAIGTNSVGVAVITDEYKVPKSKFKVLGNTRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRN
RICYLQEIFSNEMAKVDDSPFHRLEESFLVEEDKKHERHPIFGNIIVDEVAYHEKYPTIYHLRKLKLVDS
YLAHAHMKFRGHFLIEGDLNPDNSVDKFLIQLVQTYNQLFEEENPINASGVDAKAILSARLSKSRRLLENLIA
EKKNGLPFNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDQYADLFLLAKNLSDAILLSDILRV
NTBEITKAPLSASIMKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKNLREDLRLKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTPRIPYVVGPLARGNSRF
AWMTRKSEETITPWNFEVVDKGSASQSFIERMTNFDKNLPNEKVLPHKSHLLYEFTVYNELTKVKYVTEGMRKPAF
LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRPNASLGTYHDLKKIKDKDFLDNEENEDIL
EDIVLTTLTFEDREMIERLKYAHLFDDKVMQKLRRTYTGWGRSLRKLINGIRDKQSGKITLDFLKSDFGPFANRNF
MQLIHDDSLTFKEDIQKAQVSGGQSLHEHIANLAGSPAIKKILQTVKVVDLKVVMGRHKPENIVIEARENQTT
QKGQKNSRERMKRIEIEGKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDQVVDHIVPQSFL
KDDSIDNKVLTTRSDKNRKGSDNVSEEVVKMKNYWRQLLNAKLIQTKRFDNLTKAERGGSELDAKGFIRKQLVET
PKLTKHVAQILDSRMNTKYIDENDKLIREVKVI TLKSKLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTA
IKKY PLESEFVYGDYKVDYDKMIKSEQIEIGKATAKYFFYSNIMNFKEITLANGEIRKPLIETNGETGEIVWDKGR
DFATVRKVLSPQVNIYKTEVQGGFSKESILPKRNSDKLIARKDWDPKKYGGFDSPTVAYSVLVAKVEKGSK

-continued

KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
 VNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKHHRDKPIREQAENI
 IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD
 (single underline: HNH domain; double underline: RuvC domain)

[0021] 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. For example, in some embodiments, a dCas9 domain comprises D10A and/or H820A mutation. dCas9 (D10A and H840A):

[0023] In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid of a Cas9 protein, e.g., one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only a fragment thereof. For example, in some embodiments, a Cas9 fusion

(SEQ ID NO: 34)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKKFKVLGNTRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRN
 RICYLQEIFSNEMAKVDDSDFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSITDKADLRLLI
 YLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLLENLIAQLPG
 EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV
 NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
 TEELLVLKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTRIPYVVGPLARGNSRF
 AWMTRKSEETITPWNFEVVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAF
 LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKKI IKDKDFLDNEENEDIL
 EDIVLTLTLEFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDQSGKTILDFLKSDGFANRNF
 MQLIHDDSLTFKEDIQKAQVSGQGSLSHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT
 QKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFL
 KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVET
 RQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYYHHAHDAYLNAVVG TALIKKY
 PKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEIVWDKGR
 DFATVRKVLSPQVNIIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSK
 KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
 VNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEII EQISEFSKRVLADANLDKVL SAYNKHHRDKPIREQAENI
 IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGD
 (single underline: HNH domain; double underline: RuvC domain)

[0022] In other embodiments, dCas9 variants having mutations other than D10A and H820A are provided, which e.g., result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H820, or other substitutions within the nuclease domains of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 (e.g., variants of SEQ ID NO: 34) are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to SEQ ID NO:34. In some embodiments, variants of dCas9 (e.g., variants of SEQ ID NO: 34) are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 34, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids or more.

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.

[0024] 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 torquus* (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).

[0025] The term “deaminase” refers to an enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uracil or deoxyuracil, respectively.

[0026] 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. For example, in some embodiments, an effective amount of a nuclease may refer to the amount of the nuclease that is sufficient to induce cleavage of a target site specifically bound and cleaved by the nuclease. In some embodiments, an effective amount of a recombinase may refer to the amount of the recombinase that is sufficient to induce recombination at a target site specifically bound and recombined by the recombinase. As will be appreciated by the skilled artisan, the effective amount of an agent, e.g., a nuclease, a recombinase, a hybrid protein, a fusion 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, the specific allele, genome, target site, cell, or tissue being targeted, and the agent being used.

[0027] The term “linker,” as used herein, refers to a chemical group or a molecule linking two molecules or moieties, e.g., a binding domain and a cleavage domain of a nuclease. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease and the catalytic domain of a recombinase. In some embodiments, a linker joins a dCas9 and a recombinase. Typically, the 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, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 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. Longer or shorter linkers are also contemplated.

[0028] 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)).

[0029] 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-propynylcytidine, 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).

[0030] The term “proliferative disease,” as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasia is also referred to as cancer.

[0031] The terms “protein,” “peptide,” and “polypeptide” are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modifica-

tion, etc. A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may 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. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an “amino-terminal fusion protein” or a “carboxy-terminal fusion protein,” respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (e.g., the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain or a catalytic domain of a recombinase. In some embodiments, a protein comprises a proteinaceous part, e.g., an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, e.g., a compound that can act as a nucleic acid cleavage agent. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, e.g., RNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[0032] The term “RNA-programmable nuclease,” and “RNA-guided nuclease” are used interchangeably herein and refer to a nuclease that forms a complex with (e.g., binds or associates with) one or more RNA 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). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though “gRNA” is used interchangeably to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (e.g., and directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 protein. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA, and comprises a stem-loop structure. For example, in some embodiments, domain (2) is homologous to a tracrRNA as depicted in FIG. 1E of Jinek et al., *Science* 337:816-821(2012), the entire contents of which is incorporated herein by reference. Other examples of gRNAs (e.g., those including domain 2) can be found in U.S. Provisional Patent Application, U.S. Ser. No. 61/874,682, filed Sep. 6, 2013, entitled “Switchable Cas9 Nucleases And Uses Thereof,” and U.S. Provisional Patent Application, U.S. Ser. No. 61/874,746, filed Sep. 6, 2013, entitled “Delivery System For Functional Nucleases,” the entire contents of each are hereby incorporated by reference in their entirety. In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an “extended gRNA.” For

example, an extended gRNA will, e.g., bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site, providing the sequence specificity of the nuclease:RNA complex. In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example Cas9 (Csn1) from *Streptococcus pyogenes* (see, e.g., “Complete genome sequence of an M1 strain of *Streptococcus pyogenes*,” Ferretti J. J., McShan W. M., Ajdic D. J., Savic D. J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A. N., Kenton S., Lai H. S., Lin S. P., Qian Y., Jia H. G., Najjar F. Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S. W., Roe B. A., McLaughlin R. E., *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., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., Eckert M. R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); and “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *Science* 337:816-821 (2012), the entire contents of each of which are incorporated herein by reference.

[0033] Because RNA-programmable nucleases (e.g., Cas9) use RNA:DNA hybridization to target DNA cleavage sites, these proteins are able to be targeted, in principle, to any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (e.g., to modify a genome) are known in the art (see e.g., Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823 (2013); Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* 339, 823-826 (2013); Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* 31, 227-229 (2013); Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013); Dicarlo, J. E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic acids research* (2013); Jiang, W. et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

[0034] The term “subject,” as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal. In some embodiments, the subject is genetically engineered, e.g., a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

[0035] The term “target site” refers to a sequence within a nucleic acid molecule that is deaminated by a deaminase or a fusion protein comprising a deaminase, (e.g., a dCas9-deaminase fusion protein provided herein).

[0036] The terms “treatment,” “treat,” and “treating,” refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one

or more symptoms thereof, as described herein. As used herein, the terms “treatment,” “treat,” and “treating” refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, e.g., to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[0037] Some aspects of this disclosure provide fusion proteins that comprise a Cas9 domain that binds to a guide RNA (also referred to as gRNA or sgRNA), which, in turn, binds a target nucleic acid sequence via strand hybridization; and a DNA-editing domain, for example, a deaminase domain that can deaminate a nucleobase, such as, for example, cytidine. The deamination of a nucleobase by a deaminase can lead to a point mutation at the respective residue, which is referred to herein as nucleic acid editing. Fusion proteins comprising a Cas9 variant or domain and a DNA editing domain can thus be used for the targeted editing of nucleic acid sequences. Such fusion proteins are useful 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, e.g., the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a subject. Typically, the Cas9 domain of the fusion proteins described herein does not have any nuclease activity but instead is a Cas9 fragment or a dCas9 protein or domain. Methods for the use of Cas9 fusion proteins as described herein are also provided.

[0038] Non-limiting, exemplary nuclease-inactive Cas9 domains are provided herein. One exemplary suitable nuclease-inactive Cas9 domain is the D10A/H840A Cas9 domain mutant: MDKKYSIGLAIGTNSVGWAVIT-DEYKVPSSKKFKVLGNTRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFS-NEMAKVDDSFHRLLESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTD-KADRLRIYLALAHMIKFRGHFLIEGDLNPDNS DVD-KLFIQLVQTYNQLFEENPINASGVDA-KAILSARLSKSRLENLIAQLPGEKKNGLFG NLIALSLGLTPNFKSNFLAEDAKLQL-SKDTYDDDLNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRY-DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGY AGYIDGGASQEEFYKFIKPILEKMDG-TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYFPFLKDNREKIEKILT-FRIPYYVGPLARGNSRFAMWTRKSEETITPWNFEE VVDKGASAQSFIERMTNFDKNLPNEKV-LPKHSLLEYFTVYNELTKVKYVTEGMRKPA FLS-

GEQKKAIVDLLFKTNRKVTVKQLKEDY-FKKIECFDSVEISGVEDRFNASLGTYHDL-
LTIKDKDFLDNEENEDILEDIV-
LTLTLFEDREMIEERLKYAHLFD-
DKVMKQLKRRRYTG WGRLSRKLINGIRDQSGK-
TILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVS
GQG DSLHEHIANLAGSPAICKGILQTVKV-
VDELVKVMGRHKPENIVEMARENQTTQKGQKN
SRERMKRIEEGKELGSQILKEH-
PVENTQLQNEKLYLYYLQNGRD-
MYVDQELDINRLSD YDVAIVPQSFLKDDSIDNKVL-
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLNA
KLIT QRKFDNLTKAERGGLSELDKAGFIKRQL-
VETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSCLVSDFRKDFQFYKVRIN-
NYHHAHDAYLNAVVGTAIKKYPKLESEFVYG
DYKVYDVRKMIKSEQEIGKATAKY-
FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSPQVNVKK-
TEVQTGGFSKESILPKRNSDKLIARKKDWDPKK
YGGFDSPTVAYSVLVAKVEKKGK-
SKKLKSVKELLGITIMERSSEFNPDLEAKGYKE
VKKDLIILPKYSLFELENGRKRMLASA-
GELQKGNELALPSKYVNFYLLASHYEKLLKGS PED-
NEQKQLFVEQHKHYLDEIEQISEF-
SKRVILADANLDKVL SAYNKHRRDKPIREQAENI
IHLFTLTNLGAPAAFKYFDTTIDRKRYT-
STKEVLDATLIHQSIITGLYETRIDLSQLGGD (SEQ ID
NO: 37; see, e.g., Qi et al., Repurposing CRISPR as an RNA-
guided platform for sequence-specific control of gene expres-
sion. Cell. 2013; 152(5):1173-83, the entire contents of which
are incorporated herein by reference).

[0039] Additional suitable nuclease-inactive Cas9 domains will be apparent to those of skill in the art based on this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A, 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).

Fusion Proteins Between Cas9 and Nucleic Acid Editing Enzymes or Domains

[0040] Some aspects of this disclosure provide fusion proteins comprising (i) a nuclease-inactive Cas9 enzyme or domain; and (ii) a nucleic acid-editing enzyme or domain. In some embodiments, the nucleic acid-editing enzyme or domain is a DNA-editing enzyme or domain. In some embodiments, the nucleic acid-editing enzyme possesses deaminase activity. In some embodiments, the nucleic acid-editing enzyme or domain comprises or is a deaminase domain. In some embodiments, the deaminase is a cytidine deaminase. In some embodiments, the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the deaminase is an APOBEC1 family deaminase. In some embodiments, the deaminase is an activation-induced cytidine deaminase (AID). In some embodiments, the deaminase is an ACF1/ASE deaminase. In some embodiments, the deaminase is an adenosine deaminase. In some embodiments, the deaminase is an ADAT family deaminase. Some nucleic-acid editing enzymes and domains as well as Cas9 fusion proteins includ-

ing such enzymes or domains are described in detail herein. Additional suitable nucleic acid-editing enzymes or domains will be apparent to the skilled artisan based on this disclosure.

[0041] The instant disclosure provides Cas9:nucleic acid-editing enzyme/domain fusion proteins of various configurations. In some embodiments, the nucleic acid-editing enzyme or domain is fused to the N-terminus of the Cas9 domain. In some embodiments, the nucleic acid-editing enzyme or domain is fused to the C-terminus of the Cas9 domain. In some embodiments, the Cas9 domain and the nucleic acid-editing enzyme or domain are fused via a linker. In some embodiments, the linker comprises a (GGGS)_n (SEQ ID NO: 91), a (G)_n, an (EAAAK)_n (SEQ ID NO: 5), or an (XP)_n motif, or a combination of any of these, wherein n is independently an integer between 1 and 30. In some embodiments, n is independently 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, or, if more than one linker or more than one linker motif is present, any combination thereof. Additional suitable linker motifs and linker configurations will be apparent to those of skill in the art. In some embodiments, suitable linker motifs and configurations include those described in Chen et al., Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev.* 2013; 65(10):1357-69, the entire contents of which are incorporated herein by reference. Additional suitable linker sequences will be apparent to those of skill in the art based on the instant disclosure.

[0042] In some embodiments, the general architecture of exemplary Cas9 fusion proteins provided herein comprises the structure:

[0043] [NH₂]-[nucleic acid-editing enzyme or domain]-[Cas9]-[COOH] or

[0044] [NH₂]-[Cas9]-[nucleic acid-editing enzyme or domain]-[COOH],

wherein NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein.

[0045] Additional features may be present, for example, one or more linker sequences between the NLS and the rest of the fusion protein and/or between the nucleic acid-editing enzyme or domain and the Cas9. Other exemplary features that may be present are localization sequences, such as nuclear localization sequences, 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 localization signal sequences and sequences of protein tags are provided herein, and 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, FIAsh tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

[0046] In some embodiments, the nucleic acid-editing enzyme or domain is a deaminase. For example, in some embodiments, the general architecture of exemplary Cas9 fusion proteins with a deaminase enzyme or domain comprises the structure:

[0047] [NH₂]-[NLS]-[Cas9]-[deaminase]-[COOH],

[0048] [NH₂]-[NLS]-[deaminase]-[Cas9]-[COOH],

[0049] [NH₂]-[Cas9]-[deaminase]-[COOH], or

[0050] [NH₂]-[deaminase]-[Cas9]-[COOH]

wherein NLS is a nuclear localization signal, NH₂ is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, a linker is inserted between the Cas9 and the deaminase. In some embodiments, the NLS is located C-terminal of the deaminase and/or the Cas9 domain. In some embodiments, the NLS is located between the deaminase and the Cas9 domain. Additional features, such as sequence tags, may also be present

[0051] One exemplary suitable type of nucleic acid-editing enzymes and domains are cytosine deaminases, for example, of the APOBEC family. The apolipoprotein B mRNA-editing complex (APOBEC) family of cytosine deaminase enzymes encompasses eleven proteins that serve to initiate mutagenesis in a controlled and beneficial manner.²⁹ One family member, activation-induced cytidine deaminase (AID), is responsible for the maturation of antibodies by converting cytosines in ssDNA to uracils in a transcription-dependent, strand-biased fashion.³⁰ The apolipoprotein B editing complex 3 (APOBEC3) enzyme provides protection to human cells against a certain HIV-1 strain via the deamination of cytosines in reverse-transcribed viral ssDNA.³¹ These proteins all require a Zn²⁺-coordinating motif (His-X-Glu-X₂₃₋₂₆-Pro-Cys-X₂₋₄-Cys) and bound water molecule for catalytic activity. The Glu residue acts to activate the water molecule to a zinc hydroxide for nucleophilic attack in the deamination reaction. Each family member preferentially deaminates at its own particular "hotspot", ranging from WRC (W is A or T, R is A or G) for hAID, to TTC for hAPOBEC3F.³² A recent crystal structure of the catalytic domain of APOBEC3G (FIG. 2) revealed a secondary structure comprised of a five-stranded β-sheet core flanked by six α-helices, which is believed to be conserved across the entire family.³³ The active center loops have been shown to be responsible for both ssDNA binding and in determining "hotspot" identity.³⁴ Overexpression of these enzymes has been linked to genomic instability and cancer, thus highlighting the importance of sequence-specific targeting.³⁵

[0052] Another exemplary suitable type of nucleic acid-editing enzymes and domains are adenosine deaminases. For example, an ADAT family adenosine deaminase can be fused to a Cas9 domain, e.g., a nuclease-inactive Cas9 domain, thus yielding a Cas9-ADAT fusion protein.

[0053] Some aspects of this disclosure provide a systematic series of fusions between Cas9 and deaminase enzymes, e.g., cytosine deaminase enzymes such as APOBEC enzymes, or adenosine deaminase enzymes such as ADAT enzymes, that has been generated in order to direct the enzymatic activities of these deaminases to a specific site in genomic DNA. The advantages of using Cas9 as the recognition agent are two-fold: (1) the sequence specificity of Cas9 can be easily altered by simply changing the sgRNA sequence; and (2) Cas9 binds to its target sequence by denaturing the dsDNA, resulting in a stretch of DNA that is single-stranded and therefore a viable substrate for the deaminase. Successful fusion proteins have been generated with human and mouse deaminase domains, e.g., AID domains. A variety of other fusion proteins between the catalytic domains of human and mouse AID and Cas9 are also contemplated. It will be understood that other catalytic domains, or catalytic domains from other deaminases, can also be used to generate fusion proteins with Cas9, and that the disclosure is not limited in this regard.

[0054] In some embodiments, fusion proteins of Cas9 and AID are provided. In an effort to engineer Cas9 fusion proteins to increase mutation rates in ssDNA, both mouse and human AID were tethered to gene V of filamentous phage (a nonspecific ssDNA binding protein). The resulting fusion proteins exhibited enhanced mutagenic activities compared to the wild type enzymes in a cell-based assay. This work demonstrates that the enzymatic activity of these proteins is maintained in and can be successfully targeted to genetic sequences with fusion proteins.³⁶

[0055] No crystal structure has yet been reported of Cas9 bound to its target DNA, and thus the portion of DNA that is single stranded in the Cas9-DNA complex (the size of the Cas9-DNA bubble) has not been delineated. However, it has been shown in a dCas9 system with a sgRNA specifically designed for the complex to interfere with transcription that transcriptional interference only occurs when the sgRNA binds to the non-template strand. This result suggests that certain portions of the DNA in the DNA-Cas9 complex are unguarded by Cas9, and could potentially be targeted by AID

in the fusion protein.¹⁶ Accordingly, both N-terminal and C-terminal fusions of Cas9 with a deaminase domain are useful according to aspects of this disclosure.

[0056] In some embodiments, the deaminase domain and the Cas9 domain are fused to each other via a linker. Various linker lengths and flexibilities between the deaminase domain (e.g., AID) and the Cas9 domain can be employed (e.g., ranging from very flexible linkers of the form (GGGS)_n (SEQ ID NO: 91) and (G)_n to more rigid linkers of the form (EAAAK)_n (SEQ ID NO: 5) and (XP)_n)³⁷ in order to achieve the optimal length for deaminase activity for the specific application.

[0057] Some exemplary suitable nucleic-acid editing enzymes and domains, e.g., deaminases and deaminase domains, that can be fused to Cas9 domains according to aspects of this disclosure are provided below. It will be understood that, in some embodiments, the active domain of the respective sequence can be used, e.g., the domain without a localizing signal (nuclear localizing signal, without nuclear export signal, cytoplasmic localizing signal).

Human AID:

(SEQ ID NO: 6)

MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGHLRNKNGCHV
 ELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCARHVADFLRGPNLSLRIFTARLYFC
 EDKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHRTFKAWEGHLENSVRLSR
QLRRILLPLYEVDDLDAFRTLGL
 (underline: nuclear localization signal; double underline:
 nuclear export signal)

Mouse AID:

(SEQ ID NO: 7)

MDSLLMKQKKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSCSLDFGHLRNKSGCHV
 ELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWPNLSLRIFTARLYFC
 EDKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRRTFKAWEGHLENSVRLSR
QLRRILLPLYEVDDLDAFRLMLGF
 (underline: nuclear localization signal; double underline:
 nuclear export signal)

Dog AID:

(SEQ ID NO: 8)

MDSLLMKQKFLYHFKNVRWAKGRHETLYCYVVKRRDSATSFSLDFGHLRNKSGCHV
 ELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFC
 EDKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGHLENSVRLSR
QLRRILLPLYEVDDLDAFRTLGL
 (underline: nuclear localization signal; double underline:
 nuclear export signal)

Bovine AID:

(SEQ ID NO: 9)

MDSLLKKQKFLYQFKNVRWAKGRHETLYCYVVKRRDSPTSFSDFGHLRNKAGCHV
 ELLFLRYISDWLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFC
 DKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHRTFKAWEGHLENSVRLSR
RQLRRILLPLYEVDDLDAFRTLGL
 (underline: nuclear localization signal; double underline:
 nuclear export signal)

Mouse APOBEC-3:

(SEQ ID NO: 10)

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLH
 HGVFKNKDNIIHAETICFLYWFDKVLKVLSPREEFKITWYMSWSPCFCAEQIVRFLATHHN
 LSLDIFSSRLYNVQDPETQQNLCLRLVQEGAQVAAMDLYEFKKCWKFVDNGRRFRP
 WKRLLTNFRYQDSKLQELRPCYIPVPSSSSTLSNICLTGKLPTREFCVGRMDPLSEE
 EFYSQFYNQVRVHLCYHYHRMKPYLCYQLEQFNGQAPLKGCLLSEKKGQHAETLFLDKIR
 SMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFPQKGLCSLW
 QSGILVDVMDLPQFTDCWNTFVNPKRPFWPWKGLEIISRTRQRLRRIKESWGLQDLVN
 DFGNLQLGPPMS
 (italic: nucleic acid editing domain)

Rat APOBEC-3:

(SEQ ID NO: 11)

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSLH
 HGVFKNKDNIIHAETICFLYWFDKVLKVLSPREEFKITWYMSWSPCFCAEQVLRFLATHHN

- continued

LSLDIFSSRLYNIRDPENQONLCRLVQEGAQVAAMDLYEFKKCKWKKFVDNGGRRFRPW
 KLLLTNFRYQDSKLQEILRPCYIPVPSSSSTLSNICLTGGLPETRFCEVERRVHLLSEEEF
 YSQFYNRQVRKHLCCYHGVKPYLCYQLEQFNQAPLKGCLLSEKKGQHAELFLDKIRSM
 ELSQVIITCYLTWSPCNCANQLAAFKRDRPDLLIHIYTSRLYFHWKRPFQKGLCSLWQSG
 ILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRHLRIKESWGLQDLVNDFG
 NLQLGPPMS
 (*italic: nucleic acid editing domain*)

Rhesus macaque APOBEC-3G:

(SEQ ID NO: 12)

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTCDPSGPPLDAKIFQGGKYSKAKYHP
EMRFLRWFKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYY
 FWKPDYQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPKPRNNLPKH
 TLLQATLGEILLRLHMDPGTFTSNFNNKPWVSGQHETLYCYKVERLHNDTWVPLNQHR
 GFLRNQAPNIGHFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISN
 NEHVSLCIFAARIYDDQGRYQEGRLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQ
 PWDGLDEHSQALSGRLRAI
 (*italic: nucleic acid editing domain; underline: cytoplasmic
 localization signal*)

Chimpanzee APOBEC-3G:

(SEQ ID NO: 13)

MKPHFRNPVERMYQDTFSDNFYNNRPILSHRNTVWLCYEVKTGKPSRPPLDAKIFRGQVY
 SKLYHPHMRFFHFWFSKWRKLHRDQEYEVWYISWSPTCKTRDVATFLAEDPKVTLTIF
 VARLYYFWDPDYQALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEFVN
 NLPKYIILLHIMLGEILLRHMDPPTFTSNFNNELWVRGRHETLYCYEVERLHNDTWVLL
 NQRRGFLCNQAPHKHGFLGRHAELCFLDVIIPFWKLDLHDQYRVTCFTSWSPCFSCAQEM
 AKFISNNKHVSLCIFAARIYDDQGRYQEGRLTLAKAGAKISIMTYSEFKHCWDTFVDHQ
 GCPFPQPDGLDEHSQALSGRLRAILQNQGN
 (*italic: nucleic acid editing domain; underline: cytoplasmic
 localization signal*)

Green monkey APOBEC-3G:

(SEQ ID NO: 14)

MNPQIRNMVEQMEPDIFVYFNNRPILSGRNTVWLCYEVKTGKPSRPPLDANIFQGLY
 PEAKDHPHMRFFHFWFSKWRKLHRDQEYEVWYISWSPTCKTRCANVATFLAEDPKVTLTIF
 VARLYYFWDPDYQALRILCQERGGPHATMKIIVINYNFQHCWNEFVDGQGGKPKPRK
 NLPKHYTLHATLGEILLRHMDPPTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWV
 LLNQHRGFLRNQAPDRHGFPGKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQK
 MAKFISNNKHVSLCIFAARIYDDQGRYQEGRLTLHRDGAKIAVMNYSEFEYCWDTFVD
 RQGRPFQPDGLDEHSQALSGRLRAI
 (*italic: nucleic acid editing domain; underline: cytoplasmic
 localization signal*)

Human APOBEC-3G:

(SEQ ID NO: 15)

MKPHFRNTVERMYRDTFSYNFNNRPILSRRNTVWLCYEVKTGKPSRPPLDAKIFRGQVY
 SELKYHPHMRFFHFWFSKWRKLHRDQEYEVWYISWSPTCKTRDMATFLAEDPKVTLTIF
 VARLYYFWDPDYQALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEFVN
 NLPKYIILLHIMLGEILLRHMDPPTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLL
 NQRRGFLCNQAPHKHGFLGRHAELCFLDVIIPFWKLDLHDQYRVTCFTSWSPCFSCAQEM
 AKFISNNKHVSLCIFTARIYDDQGRYQEGRLTLAEAGAKISIMTYSEFKHCWDTFVDHQ
 GCPFPQPDGLDEHSQALSGRLRAILQNQGN
 (*italic: nucleic acid editing domain; underline: cytoplasmic
 localization signal*)

Human APOBEC-3F:

(SEQ ID NO: 16)

MKPHFRNTVERMYRDTFSYNFNNRPILSRRNTVWLCYEVKTGKPSRPPLDAKIFRGQV
 YSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFSVWTPCPDCVAKLAELAEHPNVTLTIS
 AARLYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFD
 DNYAFLHRTLKEILLRNPMEMPHIYFHFKNLRKAYGRNESWLCFTMEVVKHHSVPS
 WKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVWYTSWSPCECAGEVAEFLA
 RSHNVNLTIFTARLYYFWDTDYQEGRLSLSQEGASVEIMGYKDFKYCWENFVYNDDEP
 FKPWKGLKYNFLFLDSKLQEILE
 (*italic: nucleic acid editing domain*)

Human APOBEC-3B:

(SEQ ID NO: 17)

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKGRSNLLWDTGVRGQ
 VYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFSVWTPCPDCVAKLAELAEHPNVTLTIS
 SAARLYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFD
 DENYAFHRTLKEILLRYLMDPDTFTFNFNNDPLVLRRTQYLYCYEVERLDNGTWVLM
 QHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQDPAQIYRVTFWISWSPCFSWGCAGE
 VRAFLQENTHVRIRIFAARIYDYDPLYKEALQMLRDAGAQSIMTYDEFEYCWDTFVY
 RQCGPFPQPDGLDEHSQALSGRLRAILQNQGN
 (*italic: nucleic acid editing domain*)

-continued

Human APOBEC-3C:

(SEQ ID NO: 18)

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCPTVEGIKRRSVSVSWKTGVFRN
 QVDSETHCHAERCFLSNFCDILSPNTKYQVTWYTSWSPCPCDAGEVAEFLARHSNVNLT
 FTARLYYFQYPCYQEGRLSLSQEGVAVEIMDYEDFKYCWFNFVYNDNEFPKPKWGLKT
 NFRLLKRRLRESLQ
 (*italic: nucleic acid editing domain*)

Human APOBEC-3A:

(SEQ ID NO: 19)

MEASPASGPRHLMDFHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQ
 AKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWGCAGEVRAFLQENTH
 VRLRIFAARIYDYDPLYKEALQMLRDAGAQSIIIVITYDEFKHCWDTFVDHQGCPFPQPWD
 GLDEHSQALSGLRLRAILQNQGN
 (*italic: nucleic acid editing domain*)

Human APOBEC-3H:

MALLTAETFRLQPMNKRRLRRPYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFI
 NEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQQ
 KGLRLLCGSQVPVEVMGFPKADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRL
 RIKIPGVRAQGRYMDILCDAEV (SEQ ID NO: 20)
 (*italic: nucleic acid editing domain*)

Human APOBEC-3D:

(SEQ ID NO: 21)

MNPQIRNPMERMRYDTFYDNFENEPILYGRSYTWLCYEVKIKGRSNLLWDTGVRGP
 VLPKRQSNHRQEVVFRFENHAEMCFLSWFCGNRLPANRRFQITWFWVSNPCLPCVVKVT
 KFLAHEPNVTLTISAARLYYRRDRWRVLLRLHKAGARVKIMDYEDFAYCWENFVC
 NEQPPMPWYKFPDNYASLHRTLKEILRNPMEMYPHIFYFHFKNLLKACGRNESWLC
 FTMEVTKHHSVAVFRKRGVFRNQVDPETHCHAERCFLSNFCDILSPNTNYEVTWYTSWSP
 CPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVS
 CWKNFVYSDDDEFPKPKWGLQTNFRLLKRRLREILQ
 (*italic: nucleic acid editing domain*)

Human APOBEC-1:

(SEQ ID NO: 22)

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMRSRKIVVRSSGKNTT
 NHVEVNFIEKFTSERDFHPSMSCSITWFLSWSPCWECQAIREFLSRHPGVTLVIYVARLF
 WHMDQQNRQGLRDLVNSGVTIQIMRASEYHCWRNFVNYPGDEAHWPQYPLWMM
 LYALELHCIIILSLPPCLKISRRWQNLTFPRLHLQCHYQTIPPHILLATGLIHPSVAWR

Mouse APOBEC-1:

(SEQ ID NO: 23)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHVSVRHTSQNTSN
 HVEVNFIEKFTTERRYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPVTLFIYIARLYH
 HTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPSPNEAYWPRYPHLWVKLYV
 LELYCIIILGLPPCLKILRRKQPQLTFFTTITLQCHYQRIPPHLLWATGLK

Rat APOBEC-1:

(SEQ ID NO: 24)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHVSIVVRHTSQNTNK
 HVEVNFIEKFTTERRYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
 ADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYPSPNEAHWPYPHLWVRVLYV
 ELYCIIILGLPPCLNILRRKQPQLTFFTTIALQSCHYQRLPPHLLWATGLK

Human ADAT-2:

(SEQ ID NO: 25)

MEAKAAPKPAASGACSVSAEETEKWMEEMHMAKEALENTEVPVGCMLVYNNEVVG
 KGRNEVNQTKNATRHAEMVAIDQVLDWCRQSGKSPSEVFEHTVLYVTVEPCIMCAAAL
 RLMKIPLVVYGCQNERFGGCGSVLNIAADLPNTGRPFQCI PGYRAEEAVMLKTFYKQ
 ENPNAPKSKVRKKECQKS

Mouse ADAT-2:

(SEQ ID NO: 26)

MEEKVESTTTPDGPVVSQVETEKWMEEMRMAKEALENIEVPVGCMLVYNNEVVGK
 GRNEVNQTKNATRHAEMVAIDQVLDWCHQHGQSPSTVFEHTVLYVTVEPCIMCAAAL
 RLMKIPLVVYGCQNERFGGCGSVLNIAADLPNTGRPFQCI PGYRAEEAVELLKTFYKQ
 ENPNAPKSKVRKDCQKS

Mouse ADAT-1:

(SEQ ID NO: 27)

MWTADEIAQLCYAHYNVRLPKQKPEPNREWTLLAAVVKIQASANQACDIPEKEVQVT
 KEVVSMTGTGTCIGQSKMRESGDI LND SHA EHARRSFQRYLLHQLHLAAVLKEDSIFVPGTQ
 RGLWRLRPDLFSFVFFSSHTPCGDASII PMLEFEEQPCCPVIRSWANNSPVQETENLED SKDKR
 NCEDPASPVAKMRLGTPARSLSNCAVHHGTQESGPVKPDVSSDLTKEEPDAANGIASGSFR
 VVDVYRTGAKCVPGETGDLREPGAAYHQVGLLRVKPGRGDRTCMSCS DKMARWNVLGCQ
 GALLMHFLFKPI YLSAVVIGKCPYSQAMRRALTGRCEETLVLPGRFGVQLEIQSGLLFEQ
 SRCVHRKRGDSPGRLVPCGAASWSAVPQQPLDVTANGFPQGTTKKEIGSPRARSRI SKVEL

- continued

FRSFQKLLSSIADDEQPD~~SIRVT~~KKLD~~TYQ~~EYKDAASAYQEAWGALRRIQPFASWIRNPPDYH
QFK (*italic: nucleic acid editing domain*)

Human ADAT-1:

(SEQ ID NO: 28)

MWTADEIAQLCYEHYGI~~R~~L~~P~~K~~K~~G~~K~~PEPNHEWTLLAAVVKIQSPADKACDTPDKPVQVT
KEVVSMGTG~~T~~K~~C~~IGQSKMRKNGDILND~~S~~HA~~E~~VIARRSFQ~~R~~YLLHQLQAATLKEDSIFVPGT
QKGVWKLRRDLIFVFFSHT~~P~~CGDASI~~I~~PMLEFEDQ~~P~~CCPVFRNWAHNS~~S~~VEASSNLEAPGNE
RKCEDPDS~~P~~VTKMRLE~~P~~GTAA~~E~~VTNGAAHHQ~~S~~FGKQ~~S~~GPIS~~P~~GIHSCDLTVEGLATVTRI
APGSAKVIDVYRTGAKCVPGEAGDSGKPGA~~F~~HQVGLLRV~~K~~PGRGDRTRSMSCSDKMARWN
VLGCQ~~G~~ALLMHLLEPIYLSAVVIGKCPYSQEAMQ~~R~~ALIGRCQNV~~S~~ALPKGFGVQELKILQSD
LLFEQ~~S~~RS~~A~~VQAKRADSPGRLVPCGAAISWSAVPEQ~~P~~LDVTANGFPQ~~G~~TTKKTIGSLQARSQIS
KVELFRSFQKLLSRIARDKWPHSLRVQKLD~~TYQ~~EYKEAASSYQEAWSTLRKQVFGSWIRNPPD
YHQFK (*italic: nucleic acid editing domain*)

[0058] In some embodiments, fusion proteins as provided herein comprise the full-length amino acid of a nucleic acid-editing enzyme, e.g., one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length sequence of a nucleic acid-editing enzyme, but only a fragment thereof. For example, in some embodiments, a fusion protein provided herein comprises a Cas9 domain and a fragment of a nucleic acid-editing enzyme, e.g., wherein the fragment comprises a nucleic acid-editing domain. Exemplary amino acid sequences of nucleic acid-editing domains are shown in the sequences above as italicized letters, and additional suitable sequences of such domains will be apparent to those of skill in the art.

[0059] Additional suitable nucleic-acid editing enzyme sequences, e.g., deaminase enzyme and domain sequences, that can be used according to aspects of this invention, e.g., that can be fused to a nuclease-inactive Cas9 domain, will be apparent to those of skill in the art based on this disclosure. In some embodiments, such additional enzyme sequences include deaminase enzyme or deaminase domain sequences that are 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%, or at least 99% similar to the sequences provided herein. Additional suitable Cas9 domains, variants, and sequences will also be apparent to those of skill in the art. Examples of such additional suitable Cas9 domains include, but are not limited to, D10A, 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).

[0060] Additional suitable strategies for generating fusion proteins comprising a Cas9 domain and a deaminase domain will be apparent to those of skill in the art based on this disclosure in combination with the general knowledge in the art. Suitable strategies for generating fusion proteins according to aspects of this disclosure using linkers or without the use of linkers will also be apparent to those of skill in the art in view of the instant disclosure and the knowledge in the art. For example, Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154(2):442-51, showed that C-terminal fusions of Cas9 with VP64 using 2 NLS's as a linker (SPKKRKRKVEAS, SEQ ID NO: 29), can be employed for transcriptional activation. Mali et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013; 31(9):833-8, reported that C-terminal fusions with VP64 without linker can be

employed for transcriptional activation. And Maeder et al., CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013; 10: 977-979, reported that C-terminal fusions with VP64 using a Gly₄Ser (SEQ ID NO:91) linker can be used as transcriptional activators.

Use of Cas9 DNA Editing Fusion Proteins for Correcting Disease-Associated Mutations

[0061] Some embodiments provide methods for using the Cas9 DNA editing fusion proteins provided herein. In some embodiments, the fusion protein is used to introduce a point mutation into a nucleic acid by deaminating a target nucleobase, e.g., a C residue. In some embodiments, the deamination of the target nucleobase results in the correction of a genetic defect, e.g., in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the genetic defect is associated with a disease or disorder, e.g., a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a Cas9 DNA editing fusion protein to introduce a deactivating point mutation into an oncogene (e.g., in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, e.g., a truncated protein lacking the function of the full-length protein.

[0062] In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via genome editing. The Cas9 deaminase fusion proteins provided herein can be validated for gene editing-based human therapeutics in vitro, e.g., by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins provided herein, e.g., the fusion proteins comprising a Cas9 domain and a nucleic acid deaminase domain can be used to correct any single point T->C or A->G mutation. In the first case, deamination of the mutant C back to U corrects the mutation, and in the latter case, deamination of the C that is base-paired with the mutant G, followed by a round of replication, corrects the mutation.

[0063] An exemplary disease-relevant mutation that can be corrected by the provided fusion proteins in vitro or in vivo is the H1047R (A3140G) polymorphism in the PI3KCA protein. The phosphoinositide-3-kinase, catalytic alpha subunit (PI3KCA) protein acts to phosphorylate the 3-OH group of the inositol ring of phosphatidylinositol. The PI3KCA gene

has been found to be mutated in many different carcinomas, and thus it is considered to be a potent oncogene.⁵⁰ In fact, the A3140G mutation is present in several NCI-60 cancer cell lines, such as, for example, the HCT116, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC).⁵¹

[0064] In some embodiments, a cell carrying a mutation to be corrected, e.g., a cell carrying a point mutation, e.g., an A3140G point mutation in exon 20 of the PI3KCA gene, resulting in a H1047R substitution in the PI3KCA protein, is contacted with an expression construct encoding a Cas9 deaminase fusion protein and an appropriately designed sgRNA targeting the fusion protein to the respective mutation site in the encoding PI3KCA gene. Control experiments can be performed where the sgRNAs are designed to target the fusion enzymes to non-C residues that are within the PI3KCA gene. Genomic DNA of the treated cells can be extracted, and the relevant sequence of the PI3KCA genes PCR amplified and sequenced to assess the activities of the fusion proteins in human cell culture.

[0065] It will be understood that the example of correcting point mutations in PI3KCA is provided for illustration purposes and is not meant to limit the instant disclosure. The skilled artisan will understand that the instantly disclosed DNA-editing fusion proteins can be used to correct other point mutations and mutations associated with other cancers and with diseases other than cancer including other proliferative diseases.

[0066] The successful correction of point mutations in disease-associated genes and alleles opens up new strategies for gene correction with applications in therapeutics and basic research. Site-specific single-base modification systems like the disclosed fusions of Cas9 and deaminase enzymes or domains also have applications in “reverse” gene therapy, where certain gene functions are purposely suppressed or abolished. In these cases, site-specifically mutating Trp (TGG), Gln (CAA and CAG), or Arg (CGA) residues to premature stop codons (TAA, TAG, TGA) can be used to abolish protein function in vitro, ex vivo, or in vivo.

[0067] The instant 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 Cas9 DNA editing fusion protein provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, e.g., a cancer associated with a PI3KCA point mutation as described above, an effective amount of a Cas9 deaminase fusion protein that corrects the point mutation or introduces a deactivating mutation into the disease-associated gene. In some embodiments, the disease is a proliferative disease. In some embodiments, the disease is a genetic disease. In some embodiments, the disease is a neoplastic disease. In some embodiments, the disease is a metabolic disease. In some embodiments, the disease is a lysosomal storage disease. Other diseases that can be treated by correcting a point mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

[0068] The instant disclosure provides methods for the treatment of additional diseases or disorders, e.g., diseases or disorders that are associated or caused by a point mutation that can be corrected by deaminase-mediated gene editing. Some such diseases are described herein, and additional 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. Exemplary suitable diseases and disorders are listed below. 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 be different, 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. Exemplary suitable diseases and disorders include, without limitation, cystic fibrosis (see, e.g., Schwank et al., Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell stem cell*. 2013; 13: 653-658; and Wu et. al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell*. 2013; 13: 659-662, neither of which uses a deaminase fusion protein to correct the genetic defect); phenylketonuria—e.g., phenylalanine to serine mutation at position 835 (mouse) or 240 (human) or a homologous residue in phenylalanine hydroxylase gene (T>C mutation)—see, e.g., McDonald et al., *Genomics*. 1997; 39:402-405; Bernard-Soulier syndrome (BSS)—e.g., phenylalanine to serine mutation at position 55 or a homologous residue, or cysteine to arginine at residue 24 or a homologous residue in the platelet membrane glycoprotein IX (T>C mutation)—see, e.g., Noris et al., *British Journal of Haematology*. 1997; 97: 312-320, and Ali et al., *Hematol*. 2014; 93: 381-384; epidermolytic hyperkeratosis (EHK)—e.g., leucine to proline mutation at position 160 or 161 (if counting the initiator methionine) or a homologous residue in keratin 1 (T>C mutation)—see, e.g., Chipev et al., *Cell*. 1992; 70: 821-828, see also accession number P04264 in the UNIPROT database at [www\[dot\]uniprot\[dot\]org](http://www[dot]uniprot[dot]org); chronic obstructive pulmonary disease (COPD)—e.g., leucine to proline mutation at position 54 or 55 (if counting the initiator methionine) or a homologous residue in the processed form of α_1 -antitrypsin or residue 78 in the unprocessed form or a homologous residue (T>C mutation)—see, e.g., Poller et al., *Genomics*. 1993; 17: 740-743, see also accession number P01011 in the UNIPROT database; Charcot-Marie-Toot disease type 4J—e.g., isoleucine to threonine mutation at position 41 or a homologous residue in FIG. 4 (T>C mutation)—see, e.g., Lenk et al., *PLoS Genetics*. 2011; 7: e1002104; neuroblastoma (NB)—e.g., leucine to proline mutation at position 197 or a homologous residue in Caspase-9 (T>C mutation)—see, e.g., Kundu et al., 3 *Bio-tech*. 2013, 3:225-234; von Willebrand disease (vWD)—e.g., cysteine to arginine mutation at position 509 or a homologous residue in the processed form of von Willebrand factor, or at position 1272 or a homologous residue in the unprocessed form of von Willebrand factor (T>C mutation)—see, e.g., Lavergne et al., *Br. J. Haematol*. 1992, see also accession number P04275 in the UNIPROT database; 82: 66-72; myotonia congenital—e.g., cysteine to arginine mutation at position 277 or a homologous residue in the muscle chloride channel gene CLCN1 (T>C mutation)—see, e.g., Weinberger et al., *The J. of Physiology*. 2012; 590: 3449-3464; hereditary

renal amyloidosis—e.g., stop codon to arginine mutation at position 78 or a homologous residue in the processed form of apolipoprotein AII or at position 101 or a homologous residue in the unprocessed form (T>C mutation)—see, e.g., Yazaki et al., *Kidney Int.* 2003; 64: 11-16; dilated cardiomyopathy (DCM)—e.g., tryptophan to Arginine mutation at position 148 or a homologous residue in the FOXD4 gene (T>C mutation), see, e.g., Minorette et. al., *Int. J. of Mol. Med.* 2007; 19: 369-372; hereditary lymphedema—e.g., histidine to arginine mutation at position 1035 or a homologous residue in VEGFR3 tyrosine kinase (A>G mutation), see, e.g., Irrthum et al., *Am. J. Hum. Genet.* 2000; 67: 295-301; familial Alzheimer's disease—e.g., isoleucine to valine mutation at position 143 or a homologous residue in presenilin1 (A>G mutation), see, e.g., Gallo et. al., *J. Alzheimer's disease.* 2011; 25: 425-431; Prion disease—e.g., methionine to valine mutation at position 129 or a homologous residue in prion protein (A>G mutation)—see, e.g., Lewis et. al., *J. of General Virology.* 2006; 87: 2443-2449; chronic infantile neurologic cutaneous articular syndrome (CINCA)—e.g., Tyrosine to Cysteine

domain fusion protein. In some embodiments, the guide RNA comprises a structure 5'-[guide sequence]-guuuuagagcua-gaaauagcaaguuu aaaaauaaggcuaguccguuaucaacugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 38), 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 Cas9: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 Cas9:nucleic acid-editing enzyme/domain fusion proteins to specific target sequences are provided below.

[0070] H1047R (A3140G) polymorphism in the phosphoinositide-3-kinase catalytic alpha subunit (PI3KCA or PIK3CA) (the position of the mutated nucleotide and the respective codon are underlined):

```
gatgacattgcatacattcgaaagaccctagccttagataaaactgagcaagaggtttg
D D I A Y I R K T L A L D K T E Q E A L

gagtatttcatgaacaaatgaatgatgcacgtcatggtggctggacaacaaaaatggat
E F M M K Q M N D A R H G G W T T K M D

tggatcttccacacaattaaacagcatgcattgaactgaaagataactgagaaaaatgaaa
W I F H T I K Q H A L N - K I T E K M K
(Nucleotide sequence - SEQ ID NO: 39; protein sequence -
SEQ ID NO: 40).
```

mutation at position 570 or a homologous residue in cryopyrin (A>G mutation)—see, e.g., Fujisawa et. al. *Blood.* 2007; 109: 2903-2911; and desmin-related myopathy (DRM)—e.g., arginine to glycine mutation at position 120 or a homologous residue in α B crystallin (A>G mutation)—see, e.g., Kumar et al., *J. Biol. Chem.* 1999; 274: 24137-24141. The entire contents of all references and database entries is incorporated herein by reference.

[0069] It will be apparent to those of skill in the art that in order to target a Cas9:nucleic acid-editing enzyme/domain fusion protein as disclosed herein to a target site, e.g., a site comprising a point mutation to be edited, it is typically necessary to co-express the Cas9:nucleic acid-editing enzyme/domain fusion protein together with a guide RNA, e.g., an sgRNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid-editing enzyme/

[0071] Exemplary suitable guide sequences for targeting a Cas9:nucleic acid-editing enzyme/domain fusion proteins to the mutant A3140G residue include, without limitation: 5'-aucggauctauuuugacuc-3' (SEQ ID NO: 41); 5'-ucggauucuuuuugacucg-3' (SEQ ID NO: 42); 5'-cuuagauaaaacugagcaag-3' (SEQ ID NO: 43); 5'-aucuuuuugacucguucuc-3' (SEQ ID NO: 44); 5'-uaaaacugagcaagaggcuu-3' (SEQ ID NO: 45); 5'-ugguggcuggacaacaaaa-3' (SEQ ID NO: 46); 5'-gcuggacaacaaaauggau-3' (SEQ ID NO: 47); 5'-guguuaauuugucguacgua-3' (SEQ ID NO: 48). Additional suitable guide sequences for targeting a Cas9:nucleic acid-editing enzyme/domain fusion protein to a mutant PI3KCA sequence, to any of the additional sequences provided below, or to additional mutant sequences associated with a disease will be apparent to those of skill in the art based on the instant disclosure.

[0072] Phenylketonuria phenylalanine to serine mutation at residue 240 in phenylalanine hydroxylase gene (T>C mutation) (the position of the mutated nucleotide and the respective codon are underlined):

```
aatcacatttttccacttcttgaagactgtggttccatgaagataacattccccag
N H I F P L L E K Y C G F H E D N I P Q

ctggaagacgttttctcaattcctgcagacttgactggtctccgcctccgacctgtggct
L E D V S Q F L Q T C T G S R L R P V A

ggcctgcttttctctcggtatttctgggtggcctggcctccgagcttccactgcaca
G L L S S R D F L G G L A F R V F H C T
(Nucleotide sequence - SEQ ID NO: 49; protein sequence - SEQ
ID NO: 50).
```

[0073] Bernard-Soulier syndrome (BSS)—cysteine to arginine at residue 24 in the platelet membrane glycoprotein IX (T>C mutation):

atgcctgcctggggagccctgttcctgctctgggccacagcagaggccaccaaggactgc
M P A W G A L F L L W A T A E A T K D C

cccagccaggtacctgccgcgcctggaaaccatggggctgtgggtggactgcaggggc
P S P R T C R A L E T M G L W V D C R G

cacggactcacggccctgctgcctgccggcccgacccgccaccttctgctggccaac
H G L T A L P A L P A R T R H L L L A N
(Nucleotide sequence - SEQ ID NO: 51; protein sequence - SEQ ID NO: 52).

[0074] Epidermolytic hyperkeratosis (EHK)—leucine to proline mutation at residue 161 in keratin 1 (T>C mutation):

ggttatggctcctgtctgcctcctgggtggcataacaagaagtcactatcaaccagagccct
G Y G P V C P P G G I Q E V T I N Q S P

cttcagccctcaatgtggagattgacctgagatccaaaaggtagtctcgagaaagg
L Q P L N V E I D P E I Q K V K S R E R
(Nucleotide sequence - SEQ ID NO: 53; protein sequence - SEQ ID NO: 54).

[0075] Chronic obstructive pulmonary disease (COPD)—leucine to proline mutation at residue 54 in α_1 -antitrypsin (T>C mutation):

gtctccctggctgaggatccccagggagatgctgccagaagacagatacatccaccat
V S L A E D P Q G D A A Q K T D T S H H

gatcaggatcacccaaccttcaacaagatcaccccaaccgggctgagttcgcttcagc
D Q D H P T F N K I T P N P A E F A F S

ctataccgccagctggcacacagtcacaacagcaccaatatcttcttctcccgtaggc
L Y R Q L A H Q S N S T N I F F S P V S

(Nucleotide sequence-SEQ ID NO: 55; protein sequence-SEQ ID NO: 56).

[0076] chronic obstructive pulmonary disease (COPD)—leucine to proline mutation at residue 78 in α_1 -antichymotrypsin (T>C mutation):

gcctccgccaaagctggacttcgctttcagcctgtacaagcagttagtcctgaaggccct
A S A N V D F A F S L Y K Q L V L K A P

gataagaatgtcatcttctccccaccgagcatctccaccgcttggccttcctgtctctg
D K N V I F S P P S I S T A L A F L S L

ggggcccataataccaccctgacagagattctcaaaggcctcaagttctacctcacggag
G A H N T T L T E I L K G L K F Y L T E

(Nucleotide sequence-SEQ ID NO: 89; protein sequence-SEQ ID NO: 90).

[0077] Neuroblastoma (NB)—leucine to proline mutation at residue 197 in Caspase-9 (T>C mutation):

ggccactgcctcattatcaacaatgtgaacttctgccgtgagtcgggctccgcacccgc
G H C L I I N N V N F C R E S G L R T R

actggctccaacatcgactgtgagaagttgcggcgctcgttctcctcgccgcatttcatg
T G S N I D C E K L R R R F S S P H F M

-continued

gtggaggtgaagggcgacctgactgccaaagaaaatgggtgctggcctttgctggagctggcg
V E V K G D L T A K K M V L A L L E L A

(Nucleotide sequence-SEQ ID NO: 57; protein sequence-SEQ ID NO: 58).

[0078] Charcot-Marie-Tooth disease type 4J—iso-leucine to threonine mutation at residue 41 in FIG. 4 (T>C mutation):

actagagctagatacttttctagtgtggagcaataatgcagaaacgaaatcgtgtcttg
T R A R Y F L V G S N N A E T K Y R V L

aagaatgatagaacagaaccaaagatttgggtcataattgatgacagcgtgtctatact
K T D R T E P K D L V I I D D R H V Y T

caacaagaagtaagggaaacttcttggccgcttggtatcttgaaatagaacaaagatggga
Q Q E V R E L L G R L D L G N R T K M G

(Nucleotide sequence-SEQ ID NO: 59; protein sequence-SEQ ID NO: 60).

[0079] von Willebrand disease (vWD)—cysteine to arginine mutation at residue 1272 in von Willebrand factor (T>C mutation):

acagatgcccccggtagagccccaccactctgtatgtggaggacatctcggaaccgcccgttg
T D A P V S P T T L Y V E D I S E P P L

cacgattttacggcagcaggtactggacctggctcttctgctggatggctcctccagg
H D F Y R S R L L D L V F L L D G S S R

ctgtccgaggtgagtttgaagtctgaaggcctttgtgggtggacatgatggagcggctg
L S E A E F E V L K A F V V D M M E R L

(Nucleotide sequence-SEQ ID NO: 61; protein sequence-SEQ ID NO: 62).

[0080] Myotonia congenital—cysteine to arginine mutation at position 277 in the muscle chloride channel gene CLCN1 (T>C mutation):

atctgtgctgctgtcctcagcaaattcatgtctgtgttctgcggggtatagagcagcca
I C A A V L S K F M S V F C G V Y E Q P

tactactactctgatattcctgacgggtgggctgtgctgtgggagtcggcgggtgttttggg
Y Y Y S D I L T V G C A V G V G R C F G

acaccacttggaggagtgtctatttagcatcgaggtcacctccacctactttgctgttcgg
T P L G G V L F S I E V T S T Y F A V R

(Nucleotide sequence-SEQ ID NO: 63; protein sequence-SEQ ID NO: 64).

[0081] Hereditary renal amyloidosis—stop codon to arginine mutation at residue 111 in apolipoprotein AII (T>C mutation):

tactttgaaaagtcaaaggagcagctgacacccctgatcaagaaggctggaacggaactg
Y F E K S K E Q L T P L I K K A G T E L

gttaacttcttgagctatttctgtggaacttggaaacacagcctgccaccagcgaagtgtc
V N F L S Y F V E L G T Q P A T Q R S V

cagcaccattgtcttccaacccagctggcctctagaacacccactggccagtcctagag
Q H H C L P T P A G L - N T H W P V L E

(Nucleotide sequence-SEQ ID NO: 65; protein sequence-SEQ ID NO: 66).

[0082] Dilated cardiomyopathy (DCM)—tryptophan to Arginine mutation at position 148 in the FOXD4 gene (T>C mutation):

ccgcacaagcgctcacgctcagcgcatctgcgccttcattagtacgcgcttccctac
P H K R L T L S G I C A F I S D R F P Y

taccgccgcaagttccccgcgcggcagaaacagcatccgccacaacctctcgctgaacgac
Y R R K F P A R Q N S I R H N L S L N D

tgcttcgtcaagatccccgcgagccggcgccaggcaagggaactactggagcctg
C F V K I P R E P G R P G K G N Y W S L

(Nucleotide sequence-SEQ ID NO: 67; protein sequence-SEQ ID NO: 68).

[0083] Hereditary lymphedema—histidine to arginine mutation at residue 1035 in VEGFR3 tyrosine kinase (A>G mutation):

gctgaggacctgtggctgagcccgctgaccatggaagatcttgtctgtacagcttccag
A E D L W L S P L T M E D L V C Y S F Q

gtggccagagggatggagtctctggttccgaaagtgcacccgcagagacctggctgct
V A R G M E F L A S R K C I R R D L A A

cggaacattctgtgtcggaagcgacgtggtgaagatctgtgactttggccttgcccg
R N I L L S E S D V V K I C D F G L A R

(Nucleotide sequence-SEQ ID NO: 69; protein sequence-SEQ ID NO: 70).

[0084] Familial Alzheimer's disease—iso-leucine to valine mutation at residue 143 in presenilin1 (A>G mutation):

gataccgagactgtggccagagagccctgcactcaattctgaatgctgccatcatgatc
D T E T V G Q R A L H S I L N A A I M I

agtgctcgttgtgtgtcatgactatcctcctggtggttctgtataaatacaggtgctataag
S V V V V M T I L L V V L Y K Y R C Y K

gtcatccatgcctggcttattatatcatctctattgttgcgtgtctttttttcattcatt
V I H A W L I I S S L L L L F F F S F I

(Nucleotide sequence-SEQ ID NO: 71; protein sequence-SEQ ID NO: 72).

[0085] Prion disease—methionine to valine mutation at residue 129 in prion protein (A>G mutation):

aagccgagtaagccaaaaaccaacatgaagcacatggctggtgctgcagcagctggggca
K P S K P K T N M K H M A G A A A A G A

gtggtggggggccttgccggtacgtgctgggaagtgccatgagcaggcccatcatacat
V V G G L G G Y V L G S A M S R P I I H

ttcggcagtgactatgaggaccgttactatcgtgaaaacatgcaccgttaccccaaccaa
F G S D Y E D R Y Y R E N M H R Y P N Q

(Nucleotide sequence-SEQ ID NO: 73; protein sequence-SEQ ID NO: 74).

[0086] Chronic infantile neurologic cutaneous articular syndrome (CINCA)—Tyrosine to Cysteine mutation at residue 570 in cryopyrin (A>G mutation):

cttcccagccgagacgtgacagtccttcttgaaaaactatggcaaattcgaaaaggggtgt
L P S R D V T V L L E N Y G K F E K G C

-continued

```
ttgatattttgtgtacgtttctcttggcctggtaaccaggagaggacctcctacttg
L I F V V R F L F G L V N Q E R T S Y L
```

(Nucleotide sequence-SEQ ID NO: 75; protein sequence-SEQ ID NO: 76).

[0087] Desmin-related myopathy (DRM)—arginine to glycine mutation at residue 120 in α B crystallin (A>G mutation):

```
gtgaagcacttctccccagaggaactcaaagttaaggtgtgggagatgtgattgaggtg
V K H F S P E E L K V K V L G D V I E V
```

```
catggaaacatgaagagcgccaggatgaacatgggttcatctccaggagttccacggg
H G K H E E R Q D E H G F I S R E F H G
```

```
aaataccggatcccgatgatgtagacctctcaccattacttcatccctgtcatctgat
K Y R I P A D V D P L T I T S S L S S D
```

(Nucleotide sequence-SEQ ID NO: 77; protein sequence-SEQ ID NO: 78).

[0088] Beta-thalassemia—one example is leucine to proline mutation at residue 115 in Hemoglobin B.

```
gagctgcactgtgacaagctgcacgtggatcctgagaacttcaggctcctgggcaacgtg
E L H C D K L H V D P E N F R L L G N V
```

```
ctggctgtgtgctggggcccatcactttggcaaagaattcacccaccagtgaggctgcc
L V C V P A H H F G K E F T P P V Q A A
```

```
tatcagaaagtgggtggctgtggctaatgccctggccacaagtatcactaagctcgc
Y Q K V V A G V A N A L A H K Y H - A R
```

(Nucleotide sequence-SEQ ID NO: 79; protein sequence-SEQ ID NO: 80).

[0089] It is to be understood that the sequences provided above are exemplary and not meant to be limiting the scope of the instant disclosure. Additional suitable sequences of point mutations that are associated with disease and amenable to correction by Cas9:nucleic acid-editing enzyme/domain fusion proteins as well as suitable guide RNA sequences will be apparent to those of skill in the art based on this disclosure.

Reporter Systems

[0090] Some aspects of this disclosure provide a reporter system that can be used for detecting deaminase activity of the fusion proteins described herein. In some embodiments, the reporter system is a luciferase-based assay in which deaminase activity leads to expression of luciferase. To minimize the impact of potential substrate promiscuity of the deaminase domain (e.g., the 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 reporter system) is minimized. In some embodiments, an intended target residue is located in an ACG mutated start codon of the luciferase gene that is unable to initiate translation. Desired deaminase activity results in a ACG>AUG modification, thus enabling translation of luciferase and detection and quantification of the deaminase activity.

[0091] In some embodiments, in order to minimize single-stranded C residues, a leader sequence is inserted between the mutated start codon and the beginning of the luciferase gene which consists of a stretch of Lys (AAA), Asn (AAT), Leu (TTA), Ile (ATT, ATA), Tyr (TAT), or Phe (TTT) residues. The

resulting mutants can be tested to ensure that the leader sequence does not adversely affect luciferase expression or activity. Background luciferase activity with the mutated start codon can be determined as well.

[0092] The reporter system can be used to test many different sgRNAs, e.g., in order to determine which residue(s) with respect to the target DNA sequence the respective deaminase (e.g., AID enzyme) will target (FIG. 3). Because the size of the Cas9-DNA bubble is not known, sgRNAs that target non-template strand can also be tested in order to assess off-target effects of a specific Cas9 deaminase fusion protein. In some embodiments, such sgRNAs are designed such that the mutated start codon will not be base-paired with the sgRNA.

[0093] Once fusion proteins that are capable of programmable site-specific C to U modifications have been identified, their activities can be further characterized. The data from the luciferase assays can, for example, be integrated into heat maps that describe which nucleotides, with respect to the sgRNA target DNA, are being targeted for deamination by a specific fusion protein. In some embodiments, the position that results in the highest activity in the luciferase assay for each fusion is considered the “target” position, while all others are considered off-target positions.

[0094] In some embodiments, Cas9 fusions with various APOBEC3 enzymes, or deaminase domains thereof, are provided. In some embodiments, Cas9 fusion proteins with other nucleic acid editing enzymes or catalytic domains are provided, including, for example, ssRNA editing enzymes, such as the cytidine deaminases APOBEC1 and ACF1/ASF, as well as the ADAT family of adenosine deaminases,³⁸ that can

be used for ssDNA editing activity when fused to Cas9. The activity of such fusion proteins can be tested using the same reporter systems and assays described above.

[0095] In some embodiments, a reporter system is provided herein that includes 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.

[0096] The description of exemplary embodiments of the reporter systems above is provided for illustration purposes

only and not meant to be limiting. Additional reporter systems, e.g., variations of the exemplary systems described in detail above, are also embraced by this disclosure.

EXAMPLES

Example 1

Fusion Proteins

[0097] Exemplary Cas9:deaminase fusion proteins are provided below:

Cas9: human AID fusion (C-terminal) (SEQ ID NO: 30)

MDSLMLNRRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPS

KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE

MAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSDTDK

ADRLRIYALAHMIKFRGHFLIEGDLNPDNSDVKLFILQVQTYNQLFEENPINASGVDA

KAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK

DTYDDDLNLLAQIGDQYADLFLAAKNLSDAIILSDILRVNTEITKAPLSASMIKRYDEH

HQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE

LLVKLNREDLLRQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPY

YVGPLARGNSRFAMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVL

KHSLLEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED

YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR

EMIEERLKYTAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDQSGKTILDFLKSDGF

ANRNFQQLIHDDSLTFKEDIQKAQVSGQDSLHEHIANLAGSPAIAKKGILQTVKVVDL

VKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQL

QNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLLKDDSIDNKVLRSDKNR

GKSDNVPSEEVVKMKNYWRQLLNAKLIITQRKFDNLTKAERGGLSELDKAGFIKRLV

ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDPRKDFQFYKVREINNYH

HAHDAYLNAVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYS

NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEV

QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGSKKLK

SVKELLGITIMERSSEKPNIDFLEAGYKEVKDLIIKLPKYSLEFLENGRKRMLASAGE

LQKGNELALPSKYVNFYLAHYEKLKGPEDNEQQLFVEQHKHYLDEIEIEISFESKR

VILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAFKYFDTTIDRKRYTST

KEVLDTLIHQSTIGLYETRIDLSQLGGDGGGGSGGGSGGGGSYVVKRRDSATSFSL

DFGYLRNKGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP

NLSLRIFTARLYFCEDRKAEPGLRRLHRAVQIAIMTFKDYFYCWNTFVENHERTFKA

WEGLEHNSVRLSRQLRRILPLLYEVDLLRDAFRTLGL

(underline: nuclear localization signal;
double underline: nuclear export signal,
bold: linker sequence)

-continued

Cas9: human AID fusion (N-terminal)

(SEQ ID NO: 31)

MDSLLMNRKKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHV
 ELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFC
 EDRKAEEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGHENSRLSR
 QLRRIILLP**GGGGS****GGGGS****GGGGS**DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLG
 NTDHRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS
 FFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSDTKADRLLIYL
 ALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
 SKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
 NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
 ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE
 DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVGPLAR
 GNSRFAWMTRKSEETIPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSHLLYE
 YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLPKTNRKVTVKQLKEDYFKKIECF
 DSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLK
 TYAHLFDDKVMKQLKRRRYTGWGLSRKLINGIRDKQSGKTILDFLKSDFANRNFQ
 LIHDDSLTFKEDIQKAQVSGQGSLSHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRH
 KPNIVIAMARENQTTQKGQKNSRERMKRIEEGIELGSQILKEHPVENTQLQNEKLYLY
 YLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLRSDKNRGKSDNVPS
 EEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH
 VAQILDSRMNTKYDENDKLIREVKVITLSKSLVSDFRKDFQFYKVINNYHHAHDAYL
 NAVVGTALIKKYPKLESEFVYGDYKVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
 EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIVKKTEVQTGGFSKE
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLSVKELLGIT
 IMERSSEFKNPIDFLEAKGYEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELA
 LPSKYVNFLYLASHYEKLKGPSPEDNEQKQLFVEQHKHYLDEIEQISEFSKRVLADANL
 DKVLSAYNKHRRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLI
 HQSITGLYETRIDLSQLGGD

(underline: nuclear localization signal; bold: linker sequence)

Cas9: mouse AID fusion (C-terminal)

(SEQ ID NO: 32)

MDSLLMNRKKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPS
 KKFVKGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
 MAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVSDTK
 ADRLLIYALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDA
 KAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
 DTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEH
 HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
 LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPY
 YVGPLARGNSRFAWMTRKSEETIPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL

-continued

KHSLLYEFYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED
 YFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
 EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGF
 ANRNFQMQLIHDDSLTFKEDIQKAQVSGQDSLHEHIANLAGSPAIIKKGILQTVKVVDEL
 VKVMGRHKPENIVIEARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFCLKDDSIDNKVLTRSDKNR
 GKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV
 ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNHYH
 HAHDAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYS
 NIMNFPKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLK
 SVKELLGITIMERSSEFKNPIDFLEAKGYEVKKDLIIKLPKYSLFELNGRKRMLASAGE
 LQKGNELALPSKYVNFYLLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKR
 VILADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVLDTLIHQSIITGLYETRIDLSQLGGDGGGGSGGGSGGGGSYVVKRRDSATSCSL
 DFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNP
 NLSLRIFTARLYFCEDRKAEPGLRRLHRAQVQIGIMTFKDYFYCWNTFVENRERTFKA
 WEG LHENS VRLTRQLRRILLPLYEVDLRLDAFRMLGF

(underline: nuclear localization signal; bod: linker sequence;
 double underline: nuclear export signal)

Cas9: human APOBEC-3G fusion (N-terminal)

(SEQ ID NO: 33)

SPKKKRKVEASMELKYHPEMRFFHWFSKWRKLHRDQEYEV TWYISWSPCTKCTRDMA
 TFLAEDPKVTLTIFVARLYYFWDPDYQEA LRSLCQKRDGPRATMKIMNYDEFQHCWSK
 FVYSQRELFPWNNLPKYYILLHIMLGEILRHSM DPPTFTFNFNNEPWRGRHETYLCYE
 VERMHN DTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRV T
 CFTSWSPCFSQAEKMAKFKSNKHVSLCIFTARIYDDQGRQEGRLTAEAGAKISIMTY
 SEFKHCWDTFVDHQGCPFPQPDGLDEHSQDLSGRRLAILQNQENS**SPKKKRKVEASSP**
KKKRKVEASKKYSIGLAIGTNSVGAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
 FDSGETAEATRLKRTARRRYTRKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKK
 HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEG
 DLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGE
 KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLNLLAQIGDQYADLFL
 AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
 DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH
 QIHLGELHAILRRQEDFYPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRKSEETI
 TPWNFEVVVKGASAQSFIERMTNFDKNLPNEKVLPHKHSLLYEFYFTVYNELTKVKYVTE
 GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
 GTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIERLKTYAHLFDDKVMKQL
 KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKA

-continued

QVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDLVKVMGRHKPENIVIMARENQTT
 QKGQKNSRERMKRIEEDIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
 DINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQL
 LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
 NDKLIREVKVITLKSCLVSDFRKDFQFYKVRINNYHHADAYLNAVVGITALIKKYPKL
 ESEFVYGDYKVDVRKMIKSEQEI GKATKAYFFYSNIMNFFKTEITLANGEIRKRPLIET
 NGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKK
 DWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEKKNPIDFLE
 AKGYKEVKKDLIIKLPKYSLEFELENGRKRMLASAGELQKGNELALPSKYVNFYLYASHY
 EKLKGSPEDEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHDKPI
 REQAENIIHLFTLTNLGAPAFKYFDTTIDRKRYTSTKEVLDTLHQSI TGLYETRIDL SQ
 LGGD
 (underline: nuclear localization signal; bold: linker (1 NLS),
 Cas9: human APOBEC-1 fusion (N-terminal) (SEQ ID NO: 92.)
SPKKKRKVEASMTSEKGPSTGDPTLRRRIEPEWFDVFDPRELRKEACLLYEIKWGMSR
 KIWRSSGKNTTNHVEVNFIIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHP
 GVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEA
 HWPQYPPLWMLYALELHCIIISLPPCLKISRRWQNHLTFRLHLQNCYQTI PHILLA
 TGLIHPSVAWR**SPKKKRKVEASSPKKKRKVEAS**DKKYSIGLAIGTNSVGWAVITDEYK
 VPSKKFKVLGNTDRHSIIKKNLIGALLPDSGETAEATRLKRTARRRYTRRKNRICYLQEIF
 SNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKLVDST
 DKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGV
 DAKAILSARLSKSRRENLI AQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQL
 SKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
 EHHQDLTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGT
 EELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRI
 PYYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASQSFIERMTNFDKNLPNEK
 VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL
 KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKI IKDKDFLDNEENEDI LEDIVLTLTLF
 EDREMIERLKYTAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKS
 DGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVD
 DELVKVMGRHKPENIVIMARENQTTQKGQKNSRERMKRIEEDIKELGSQILKEHPVEN
 TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSD
 KNRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
 QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVRIN
 NYHHADAYLNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEI GKATKAYF
 FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKK
 TEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGSK
 KLKSVKELLGITIMERSSEKKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFELENGRKRMLAS

-continued

AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEF
SKRVILADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAFKYPDTTIDRKRY
TSTKEVL DATLIHQSI TGLYETRIDLSQLGGD
(underline: nuclear localization signal; bold: linker (1 NLS),
Cas9: human ADAT1 fusion (N-terminal) (SEQ ID NO: 35)
MDSLMLNRRKFLYQFKNVRWAKGRRETYLCSMGTGTCIGQSKMRKNGDILND SHA E VI
ARRSFQRYLLHQ LQLAATLKEDSIFVPGTQKGVWKLRRDLIFVFFSSHTPCGDASIIPMLEFED
QPCCPVFRNWAHNSSVEASSNLEAPGNERKCEDPDSVPTKKMRLEPGTAAREVTNGAAHHQ
SFGKQKSGPISPGIHS CDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSGKPGAAFH
QVGLLRVKPGRGRDRTRMS CSDKMARWNVLGCQGALLMHLLEPIYLSAVVIGKCPYSQEA
MQRALIGRCQNV SALPKGFGVQELKILQSDLLFEQSRSAVQAKRADSPGRLVPCGA AISWSAV
PEQPLDVTANGFPQGT TTKTIGSLQARSQISKVELFRSFQKLLSRIARDKWP HSLRVQKLD TY
QEYKEAASSYQEA WSTLRKQVFGSWIRNPPDYHQFGGGGSGGGGSGGGGSDKKYSIGLAI
GTNSVGWAVITDEYKVP SKKFKVLGNTDRHSI KKNLIGALLFD SGETAEATRLKRTARR
RYTRRKNRICY LQEIFSNEMAKVDDSPFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKPRGHFLIEGDLNPDNSDV DKLFIQLVQ
TYNQLFEE NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQ LSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ
EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY
PFLKDNREKIEKILTFRIPYVVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASQS
FIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIV
DLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL LKIKDKDFLDN
EENEDILEDIVLTLT LTFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN
GIRDKQSGKTILDFLKS DGFANRNFMLIHDDSLTFKEDIQKAVSGQGDSLHEHIANLA
GSPA I KKGILQT VKVVD ELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEG
IKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF
LKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVG TALI KKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSINIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV
RKVLSMPQVNI VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYS
VLVVAKEVGKSKKLKSVKELLGITEVIERSSF EKNPIDFLEAKGYKEVKKDLIIKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVE

-continued

QHKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAP
 AAFKYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGD
 (underline: nuclear localization signal; bold: linker sequence)
 Cas9: human ADAT1 fusion (-terminal) (SEQ ID NO: 36)
MDSLLMNRKFLYQFKNVRWAKGRRETYLCDKKYSIGLAIGTNSVGWAVITDEYKVPS
 KKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRKNRICYLQEIFSNE
 MAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS~~TDK~~
 ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV~~DKLFIQ~~LVQTYNQLFEENPINASGVDA
 KAILSARLSKSRRENLI~~QAQLPGEK~~KNGLFGNLI~~ALSLGL~~TPNPKSNFDLAEDAKLQLSK
 DTYDDDLNLLAQIGDQYADLFLAAKNLSDAI~~LLSDIL~~RVNTEITKAPLSASMIKRYDEH
 HQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
 LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTPRIPY
 YVGPLARGNSRFAMWTRKSEETITPWNFEVV~~DKGAS~~AQSFIERMTNFDKNLPNEKVL~~P~~
 KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS~~GEQK~~KAIVDLLFKTNRKVTVKQLKED
 YFKKIECFDSVEISGVEDRFNASLGT~~YHDL~~LKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
 EMIEERLKYTAHLFDDKVMKQLKRRRYTGWRLSRKLINGIRDKQSGKTILD~~FLKSD~~GF
 ANRNFMLIHDDSLTFEKDIQKAQVSGQDSLHEHIANLAGSPA~~IKKGI~~LQTVKVVDEL
 VKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE~~E~~GKELGSQILKEHPVENTQL
 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQ~~SFLK~~DDSIDNKVLT~~RS~~DKNR
 GKS~~DN~~VPSEEVVKKMKNYWRQLLN~~AKLIT~~QRKFDNLTKAERGGLSEL~~D~~KAGFIK~~RQ~~LV
 ETRQITKHVAQIILDSRMNTKYDENDKLIREVKVITLKS~~KL~~VSDFRKDFQFYK~~VRE~~INNYH
 HAHDAYLNAVVG~~TALIK~~KYPKLESEFVYGDYKVYDVRKMI~~AKSE~~QEIGKATAKYFFYS
 NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK~~VLSMPQ~~VNIVKKTEV
 QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGPDSPTVAYS~~VLV~~AKVEKGKSK~~KLK~~
 SVKELLGITIMERS~~SFEKN~~PIDFLEAKGYEVKKDLIIKLPKYSLFELENG~~RK~~RLASAGE
 LQKGNELALPSKYVNF~~LYLASH~~EK~~LK~~GPS~~EDNE~~QQLFVEQHKHYLDEIIEQISEFSKR
 VILADANLDKVL SAYNKH RDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTST
 KEVLDATLIHQ SITGLYETRIDLSQLGGD**GGGGSGGGGS**SMGTGTCIGQSKMRKNGDIL
 ND~~SHAE~~VIARRSFQRYLLHQ~~LQ~~LAATLKEDSIFVPGTQKG~~VWKL~~RRLD~~LI~~FVFFSSHTPCGDASI
 I~~PMLE~~FEDQCCPVFRNWAHNSSVEASSNLEAPGNERKCEDPDS~~PVT~~KMRLEPGTAAREVT
 NGA~~AHH~~QSFQKQSGPISPGIHSCDLTVEGLATVTRIAPGSAKVIDVYRTGAKCVPGEAGDSG
 KPGAAFHQVGLLRVKPGRGDRTRSMSCSDKMARWNVLGCQ~~GALLMHL~~LEEPIYLSAVVIGK
 CPYSQ~~EAMQ~~RALIGRCQNV~~SAL~~PKGFGVQELKILQSDLLFEQSRSAVQA~~KRAD~~SPGRLVPCGA
 AISWSAVPEQPLDVTANGFPQGT~~TKKTIG~~SLQARSQISKVELFRSFQKLLSRIARDK~~WPH~~SLRV
 QKLD~~TYQEY~~KEAASSYQEAWS~~TLRKQ~~VFGSWIRNPPDYHQF
 (underline: nuclear localization signal; bold: linker sequence)

Example 2

Correction of a PI3K Point Mutation by a Cas9 Fusion Protein

[0098] An A3140G point mutation in exon 20 of the PI3KCA gene, resulting in an H1047R amino acid substitution in the PI3K protein is corrected by contacting a nucleic acid encoding the mutant protein with a Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the encoding PI3KCA gene. The A3140G point mutation is confirmed via genomic PCR of the respective exon 20 sequence, e.g., generation of a PCR amplicon of nucleotides 3000-3250, and subsequent sequencing of the PCT amplicon.

[0099] Cells expressing a mutant PI3K protein comprising an A3140G point mutation in exon 20 are contacted with an expression construct encoding the Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the antisense strand of the encoding PI3KCA gene. The sgRNA is of the sequence 5'-aucggauctauuuugacucguuuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu-3' (SEQ ID NO: 81); 5'-ucggaucuauuuugacucgguuuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu-3' (SEQ ID NO: 82); 5'-cuuagauaaaacugagcaagguuuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu-3' (SEQ ID NO: 83); 5'-aucuaauuuugacucguuucguuuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu-3' (SEQ ID NO: 84); 5'-uaaaacugagcaagagguuuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu-3' (SEQ ID NO: 85); 5'-ugguggcuggacacaaaaaguuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu-3' (SEQ ID NO: 87); or 5'-guguuaauuugucguacguuuuagagcuagaaaagcaaguuaaauaaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcguuuuu (SEQ ID NO: 88).

The cytosine deaminase activity of the Cas9:AID or the Cas9:APOBEC1 fusion protein results in deamination of the cytosine that is base-paired with the mutant G3140 to uridine. After one round of replication, the wild type A3140 is restored. Genomic DNA of the treated cells is extracted and a PCR amplicon of nucleotides 3000-3250 is amplified with suitable PCR primers. The correction of the A3140G point mutation after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 3

Correction of a Presenilin 1 Point Mutation by a Cas9 Fusion Protein

[0100] An A->G point mutation in codon 143 of the presenilin1 (PSEN1) gene, resulting in an I143V amino acid substitution in the PSEN1 protein is corrected by contacting a nucleic acid encoding the mutant PSEN1 protein with a Cas9:

AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the encoding PSEN1 gene. See, e.g., Gallo et. al., *J. Alzheimer's disease*. 2011; 25: 425-431 for a description of an exemplary PSEN1 I143V mutation associated with familial Alzheimer's Disease. The A->G point mutation is confirmed via genomic PCR of the respective PSEN1 sequence, e.g., generation of a PCR amplicon of about 100-250 nucleotides around exon 143, and subsequent sequencing of the PCT amplicon.

[0101] Cells expressing the mutant PSEN1 protein are contacted with an expression construct encoding the Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the antisense strand of the encoding PSEN1 gene. The cytosine deaminase activity of the Cas9:AID or the Cas9:APOBEC1 fusion protein results in deamination of the cytosine that is base-paired with the mutant G in codon 143 to uridine. After one round of replication, the wild type A is restored. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the A->G point mutation after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 4

Correction of an α_1 -Antitrypsin Point Mutation by a Cas9 Fusion Protein

[0102] A T->C point mutation in codon 55 of the α_1 -antitrypsin gene, resulting in an L55P amino acid substitution in the α_1 -antitrypsin protein is corrected by contacting a nucleic acid encoding the mutant α_1 -antitrypsin protein with a Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the encoding α_1 -antitrypsin gene. See, e.g., Poller et al., *Genomics*. 1993; 17: 740-743 for a more detailed description of an exemplary codon 55 T->C mutation associated with chronic obstructive pulmonary disease (COPD). The T->C point mutation is confirmed via genomic PCR of the respective α_1 -antitrypsin sequence encoding codon 55, e.g., generation of a PCR amplicon of about 100-250 nucleotides, and subsequent sequencing of the PCT amplicon.

[0103] Cells expressing the mutant α_1 -antitrypsin protein are contacted with an expression construct encoding the Cas9:AID (SEQ ID NO: 30) or a Cas9:APOBEC1 (SEQ ID NO: 92) fusion protein and an appropriately designed sgRNA targeting the fusion protein to the mutated nucleotide in codon 55 on the sense strand in the encoding α_1 -antitrypsin gene. The cytosine deaminase activity of the Cas9:ADAT1 fusion protein results in deamination of the mutant cytosine to uridine thus correcting the mutation. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the A->G point mutation in codon 55 of the α_1 -antitrypsin gene after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon

Example 5

Correction of a Von Willebrand Factor Point Mutation by a Cas9 Fusion Protein

[0104] A T->C point mutation in codon 509 of the von Willebrand factor gene, resulting in a C509A amino acid

substitution in the von Willebrand factor protein is corrected by contacting a nucleic acid encoding the mutant von Willebrand factor protein with a Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding von Willebrand factor gene. See, e.g., Lavergne et al., *Br. J. Haematol.* 1992; 82: 66-7, for a description of an exemplary von Willebrand factor C509A mutation associated with von Willebrand disease (vWD). The T->C point mutation is confirmed via genomic PCR of the respective von Willebrand factor genomic sequence, e.g., generation of a PCR amplicon of about 100-250 nucleotides around exon 509, and subsequent sequencing of the PCT amplicon.

[0105] Cells expressing the mutant von Willebrand factor protein are contacted with an expression construct encoding the Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding von Willebrand factor gene. The cytosine deaminase activity of the Cas9:ADAT1 fusion protein results in deamination of the mutant cytosine in codon 509 to uridine, thus correcting the mutation. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the T->C point mutation in codon 509 of the von Willebrand factor gene after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

Example 6

Correction of a Caspase 9 Point Mutation by a Cas9 Fusion Protein—Neuroblastoma

[0106] A T->C point mutation in codon 197 of the Caspase-9 gene, resulting in an L197P amino acid substitution in the Caspase-9 protein is corrected by contacting a nucleic acid encoding the mutant Caspase-9 protein with a Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding Caspase-9 gene. See, e.g., Lenk et al., *PLoS Genetics*. 2011; 7: e1002104, for a description of an exemplary Caspase-9 L197P mutation associated with neuroblastoma (NB). The T->C point mutation is confirmed via genomic PCR of the respective Caspase-9 genomic sequence, e.g., generation of a PCR amplicon of about 100-250 nucleotides around exon 197, and subsequent sequencing of the PCT amplicon.

[0107] Cells expressing the mutant Caspase-9 protein are contacted with an expression construct encoding the Cas9:ADAT1 fusion protein (SEQ ID NO: 35 or 36) and an appropriately designed sgRNA targeting the fusion protein to the mutation site in the sense strand of the encoding Caspase-9 gene. The cytosine deaminase activity of the Cas9:ADAT1 fusion protein results in deamination of the mutant cytosine in codon 197 to uridine, thus correcting the mutation. Genomic DNA of the treated cells is extracted and a PCR amplicon of 100-250 nucleotides is amplified with suitable PCR primers. The correction of the T->C point mutation in codon 197 of the Caspase-9 gene after treatment of the cells with the fusion protein is confirmed by sequencing the PCR amplicon.

REFERENCES

[0108] 1. Humbert O, Davis L, Maizels N. Targeted gene therapies: tools, applications, optimization. *Crit Rev Biochem Mol.* 2012; 47(3):264-81. PMID: 22530743.

- [0109] 2. Perez-Pinera P, Ousterout D G, Gersbach C A. Advances in targeted genome editing. *Curr Opin Chem Biol.* 2012; 16(3-4):268-77. PMID: 22819644.
- [0110] 3. Urnov F D, Rebar E J, Holmes M C, Zhang H S, Gregory P D. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* 2010; 11(9):636-46. PMID: 20717154.
- [0111] 4. Joung J K, Sander J D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol.* 2013; 14(1):49-55. PMID: 23169466.
- [0112] 5. Charpentier E, Doudna J A. Biotechnology: Rewriting a genome. *Nature.* 2013; 495, (7439):50-1. PMID: 23467164.
- [0113] 6. Pan Y, Xia L, Li A S, Zhang X, Sirois P, Zhang J, Li K. Biological and biomedical applications of engineered nucleases. *Mol Biotechnol.* 2013; 55(1):54-62. PMID: 23089945.
- [0114] 7. De Souza, N. Primer: genome editing with engineered nucleases. *Nat Methods.* 2012; 9(1):27. PMID: 22312638.
- [0115] 8. Santiago Y, Chan E, Liu P Q, Orlando S, Zhang L, Urnov F D, Holmes M C, Guschin D, Waite A, Miller J C, Rebar E J, Gregory P D, Klug A, Collingwood T N. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA.* 2008; 105(15):5809-14. PMID: 18359850.
- [0116] 9. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane C R, Lim E P, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley G Q, Lander E S. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* 1999; 22(3):231-8. PMID: 10391209.
- [0117] 10. Jansen R, van Embden J D, Gaastra W, Schouls L M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.* 2002; 43(6):1565-75. PMID: 11952905.
- [0118] 11. Mali P, Esvelt K M, Church G M. Cas9 as a versatile tool for engineering biology. *Nat Methods.* 2013; 10(10):957-63. PMID: 24076990.
- [0119] 12. Jore M M, Lundgren M, van Duijn E, Bultema J B, Westra E R, Waghmare S P, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer M R, Barendregt A, Shou K, Snijders A P, Dickman M J, Doudna J A, Boekema E J, Heck A J, van der Oost J, Brouns S J. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol.* 2011; 18(5):529-36. PMID: 21460843.
- [0120] 13. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science.* 2010; 327(5962):167-70. PMID: 20056882.
- [0121] 14. Wiedenheft B, Sternberg S H, Doudna J A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature.* 2012; 482(7385):331-8. PMID: 22337052.
- [0122] 15. Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* 2013; 21(11):562-7. PMID: 24095303.
- [0123] 16. Qi L S, Larson M H, Gilbert L A, Doudna J A, Weissman J S, Arkin A P, Lim W A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013; 152(5):1173-83. PMID: 23452860.
- [0124] 17. Perez-Pinera P, Kocak D D, Vockley C M, Adler A F, Kabadi A M, Polstein L R, Thakore P I, Glass K A,

- Ousterout D G, Leong K W, Guilak F, Crawford G E, Reddy T E, Gersbach C A. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013; 10(10):973-6. PMID: 23892895.
- [0125] 18. Mali P, Aach J, Stranges P B, Esvelt K M, Moosburner M, Kosuri S, Yang L, Church G M. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013; 31(9):833-8. PMID: 23907171.
- [0126] 19. Gilbert L A, Larson M H, Morsut L, Liu Z, Brar G A, Torres S E, Stern-Ginossar N, Brandman O, Whitehead E H, Doudna J A, Lim W A, Weissman J S, Qi L S. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154(2):442-51. PMID: 23849981.
- [0127] 20. Larson M H, Gilbert L A, Wang X, Lim W A, Weissman J S, Qi L S. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. 2013; 8(11):2180-96. PMID: 24136345.
- [0128] 21. Mali P, Yang L, Esvelt K M, Aach J, Guell M, DiCarlo J E, Norville J E, Church G M. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339(6121):823-6. PMID: 23287722.
- [0129] 22. Cole-Strauss A, Yoon K, Xiang Y, Byrne B C, Rice M C, Gryn J, Holloman W K, Kmiec E B. Correction of the mutation responsible for sickle cell anemia by an RNA-DNA oligonucleotide. *Science*. 1996; 273(5280):1386-9. PMID: 8703073.
- [0130] 23. Tagalakis A D, Owen J S, Simons J P. Lack of RNA-DNA oligonucleotide (chimeraplast) mutagenic activity in mouse embryos. *Mol Reprod Dev*. 2005; 71(2):140-4. PMID: 15791601.
- [0131] 24. Ray A, Langer M. Homologous recombination: ends as the means. *Trends Plant Sci*. 2002; 7(10):435-40. PMID 12399177.
- [0132] 25. Britt A B, May G D. Re-engineering plant gene targeting. *Trends Plant Sci*. 2003; 8(2):90-5. PMID: 12597876.
- [0133] 26. Vagner V, Ehrlich S D. Efficiency of homologous DNA recombination varies along the *Bacillus subtilis* chromosome. *J Bacteriol*. 1988; 170(9):3978-82. PMID: 3137211.
- [0134] 27. Saleh-Gohari N, Helleday T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res*. 2004; 32(12):3683-8. PMID: 15252152.
- [0135] 28. Lombardo A, Genovese P, Beausejour C M, Colleoni S, Lee Y L, Kim K A, Ando D, Urnov F D, Galli C, Gregory P D, Holmes M C, Naldini L. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol*. 2007; 25(11):1298-306. PMID: 17965707.
- [0136] 29. Conticello S G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol*. 2008; 9(6):229. PMID: 18598372.
- [0137] 30. Reynaud C A, Aoufouchi S, Faili A, Weill J C. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nat Immunol*. 2003; 4(7):631-8.
- [0138] 31. Bhagwat A S. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst)*. 2004; 3(1):85-9. PMID: 14697763.
- [0139] 32. Navaratnam N, Sarwar R. An overview of cytidine deaminases. *Int J Hematol*. 2006; 83(3):195-200. PMID: 16720547.
- [0140] 33. Holden L G, Prochnow C, Chang Y P, Bransteitter R, Chelico L, Sen U, Stevens R C, Goodman M F, Chen X S. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature*. 2008; 456(7218):121-4. PMID: 18849968.
- [0141] 34. Chelico L, Pham P, Petruska J, Goodman M F. Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem*. 2009; 284(41): 27761-5. PMID: 19684020.
- [0142] 35. Pham P, Bransteitter R, Goodman M F. Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry*. 2005; 44(8):2703-15. PMID 15723516.
- [0143] 36. Barbas C F, Kim D H. Cytidine deaminase fusions and related methods. *PCT Int Appl*. 2010; WO 2010132092 A2 20101118.
- [0144] 37. Chen X, Zaro J L, Shen W C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev*. 2013; 65(10):1357-69. PMID: 23026637.
- [0145] 38. Gerber A P, Keller W. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci*. 2001; 26(6):376-84. PMID: 11406411.
- [0146] 39. Yuan L, Kurek I, English J, Keenan R. Laboratory-directed protein evolution. *Microbiol Mol Biol Rev*. 2005; 69(3):373-92. PMID: 16148303.
- [0147] 40. Cobb R E, Sun N, Zhao H. Directed evolution as a powerful synthetic biology tool. *Methods*. 2013; 60(1): 81-90. PMID: 22465795.
- [0148] 41. Bershtein S, Tawfik D S. Advances in laboratory evolution of enzymes. *Curr Opin Chem Biol*. 2008; 12(2): 151-8. PMID: 18284924.
- [0149] 42. Hida K, Hanes J, Ostermeier M. Directed evolution for drug and nucleic acid delivery. *Adv Drug Deliv Rev*. 2007; 59(15):1562-78. PMID: 17933418.
- [0150] 43. Esvelt K M, Carlson J C, Liu D R. A system for the continuous directed evolution of biomolecules. *Nature*. 2011; 472(7344):499-503. PMID: 21478873.
- [0151] 44. Husimi Y. Selection and evolution of bacteriophages in cellstat. *Adv Biophys*. 1989; 25:1-43. PMID: 2696338.
- [0152] 45. Riechmann L, Holliger P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell*. 1997; 90(2):351-60. PMID: 9244308.
- [0153] 46. Nelson F K, Friedman S M, Smith G P. Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. *Virology*. 1981; 108(2):338-50. PMID: 6258292.
- [0154] 47. Rakonjac J, Model P. Roles of pIII in filamentous phage assembly. *J Mol Biol*. 1998; 282(1):25-41.
- [0155] 48. Smith G P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*. 1985; 228(4705):1315-7. PMID: 4001944.
- [0156] 49. Sheridan C. Gene therapy finds its niche. *Nat Biotechnol*. 2011; 29(2):121-8. PMID: 21301435.
- [0157] 50. Lee J W, Soung Y H, Kim S Y, Lee H W, Park W S, Nam S W, Kim S H, Lee J Y, Yoo N J, Lee S H. PIK3CA

gene is frequently mutated in breast carcinomas and hepatocellular carcinomas. *Oncogene*. 2005; 24(8):1477-80. PMID: 15608678.

[0158] 51. Ikediobi O N, Davies H, Bignell G, Edkins S, Stevens C, O'Meara S, Santarius T, Avis T, Barthorpe S, Brackenbury L, Buck G, Butler A, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Hunter C, Jenkinson A, Jones D, Kosmidou V, Lugg R, Menzies A, Mironenko T, Parker A, Perry J, Raine K, Richardson D, Shepherd R, Small A, Smith R, Solomon H, Stephens P, Teague J, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Reinhold W, Weinstein J N, Stratton M R, Futreal P A, Wooster R. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther*. 2006; 5(11):2606-12. PMID: 17088437.

[0159] All publications, patents, patent applications, publication, and database entries (e.g., sequence database entries) mentioned herein, e.g., in the Background, Summary, Detailed Description, Examples, and/or References sections, are hereby incorporated by reference in their entirety as if each individual publication, patent, patent application, publication, and database entry was specifically and individually incorporated herein by reference. In case of conflict, the present application, including any definitions herein, will control.

EQUIVALENTS AND SCOPE

[0160] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the embodiments described herein. The scope of the present disclosure is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[0161] Articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between two or more members of a group are considered satisfied if one, more than one, or all of the group members are present, unless indicated to the contrary or otherwise evident from the context. The disclosure of a group that includes “or” between two or more group members provides embodiments in which exactly one member of the group is present, embodiments in which more than one members of the group are present, and embodiments in which all of the group members are present. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[0162] It is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitation, element, clause, or descriptive term, from one or more of the claims or from one or more relevant portion of the description, is introduced into another claim. For

example, a claim that is dependent on another claim can be modified to include one or more of the limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of making or using the composition according to any of the methods of making or using disclosed herein or according to methods known in the art, if any, are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[0163] Where elements are presented as lists, e.g., in Markush group format, it is to be understood that every possible subgroup of the elements is also disclosed, and that any element or subgroup of elements can be removed from the group. It is also noted that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where an embodiment, product, or method is referred to as comprising particular elements, features, or steps, embodiments, products, or methods that consist, or consist essentially of, such elements, features, or steps, are provided as well. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[0164] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in some embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. For purposes of brevity, the values in each range have not been individually spelled out herein, but it will be understood that each of these values is provided herein and may be specifically claimed or disclaimed. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[0165] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 92

<210> SEQ ID NO 1

<211> LENGTH: 4104

<212> TYPE: DNA

<213> ORGANISM: *Streptococcus pyogenes*

-continued

<400> SEQUENCE: 1

atggataaga aataactcaat aggccttagat atcggcacaa atagcgctcgg atgggcggtg	60
atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc	120
cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa	180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt	240
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatatgtt ctttcacga	300
cttgaagagt ctttttttgt ggaagaagac aagaagcatg aacgtcatcc tatttttgga	360
aatatagtag atgaagtgtc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa	420
aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc	480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat	540
gtggacaaac tttttatcca gttggtacaa atctacaatc aattatttga agaaaaccct	600
attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga	660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat	720
ctcattgctt tgctattggg attgaccctc aattttaaat caaatttga tttggcagaa	780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg	840
caaattggag atcaatatgc tgatttgtt ttggcagcta agaatttctc agatgctatt	900
ttactttcag atatcctaag agtaaatagt gaaataacta aggcctccct atcagcttca	960
atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagtctga	1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cgatatgca	1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta	1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc	1200
aagcaacgga cctttgacaa cggctctatt ccccatcaa ttcacttggg tgagctgcat	1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt	1320
gaaaaaatct tgacttttct aattccttat tatgttggtc cattggcgcg tggcaatagt	1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa	1440
gttgctgata aagggtcttc agctcaatca tttattgaac gcattgacaaa ctttgataaa	1500
aatcttccaa atgaaaaagt actacccaaa catagtgtgc tttatgagta ttttacgggt	1560
tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt	1620
tcagggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc	1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt	1740
tcaggagtgt aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt	1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt	1860
ttaacattga cttattttga agataggggg atgattgagg aaagacttaa aacatatgct	1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga	1980
cgtttgtctc gaaaattgat taatgggtatt agggataagc aatctggcaa aacaatatta	2040
gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat	2100
agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttaa	2160
catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggat tttacagact	2220

-continued

gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcggtatt	2280
gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg	2340
aaacgaatcg aagaaggat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt	2400
gaaaatactc aattgcaaaa tgaagagctc tatctctatt atctacaaaa tggaagagac	2460
atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt	2520
gttcacaaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat	2580
aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac	2640
tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg	2700
aaagctgaac gtggagggtt gagtgaactt gataaagctg gttttatcaa acgccaatg	2760
gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact	2820
aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa	2880
ttagtttctg acttcgaaa agatttccaa ttctataaag tacgtgagat taacaattac	2940
catcatgcc atgatgcgta tctaaatgcc gtcgttgga ctgctttgat taagaaatat	3000
ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg	3060
attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactcta	3120
atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct	3180
ctaactgaaa ctaatgggga aactggagaa attgtctggg ataagggcg agattttgcc	3240
acagtgcgca aagtattgtc catgcccga gtcaatattg tcaagaaaac agaagtacag	3300
acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct	3360
cgtaaaaaag actgggatcc aaaaaatat ggtgggtttg atagtccaac ggtagcttat	3420
tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa	3480
gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaatcc gattgacttt	3540
ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat	3600
agtctttttg agttagaaaa cggtcgttaa cggatgctgg ctagtgccgg agaattacaa	3660
aaaggaaatg agctggctct gccaaagcaa tatgtgaatt ttttatatt agctagtc	3720
tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag	3780
cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt	3840
ttagcagatg ccaatttaga taaagttctt agtgcatata acaacatag agacaaacca	3900
atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc	3960
gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc taaaaagaa	4020
gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat	4080
ttgagtcagc taggaggtga ctga	4104

<210> SEQ ID NO 2

<211> LENGTH: 1367

<212> TYPE: PRT

<213> ORGANISM: Streptococcus pyogenes

<400> SEQUENCE: 2

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
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

-continued

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			

-continued

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			

-continued

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
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
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				

-continued

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> SEQ ID NO 3

<211> LENGTH: 4212

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 3

```

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc   60
ataaccgatg aatacaaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt   120
cattcgatta aaaagaatct tatcgggtgcc ctctatttcg atagtggcga aacggcagag   180
gcgactcgcc tgaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt   240
tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt   300
ttggaagagt ctttccttgt cgaagaggac aagaacatg aacggcacc cttctttgga   360
aacatagtag atgaggtggc atatcatgaa aagtaccaa cgatttatca cctcagaaaa   420
aagctagtgt actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat   480
atgataaagt tccgtgggca ctttctcatt gaggtgatc taaatccgga caactcggat   540
gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct   600
ataaatgcaa gtggcgtgga tgcaaggct attcttagcg cccgcctctc taaatccga   660
cggttagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac   720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa   780
gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca   840
caaatgggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc   900
ctcctatctg acatactgag agttaatact gagattacca aggcgcggtt atccgcttca   960
atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc ctagtccgt   1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca   1080
ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta   1140

```

-continued

gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcggaaga tctactgcga	1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat	1260
gctataactta gaaggcagga ggatttttat cggttcctca aagacaatcg tgaagagatt	1320
gagaaaaatcc taacctttcg cataccttac tatgtgggac ccttgcccg agggaaactct	1380
cggttcgcatt ggatgacaag aaagtcgaa gaaacgatta ctccatggaa ttttgaggaa	1440
gtgtgcgata aagggtcgcg agctcaatcg ttcacgaga ggatgaccaa ctttgacaag	1500
aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagt	1560
tacaatgaac tcacgaaagt taagtatgac actgagggca tgcgtaaac cgcctttcta	1620
agcggagaa acgaaagc aatagtagat ctgttattca agaccaaccg caaagtgcga	1680
gttaagcaat tgaagagga ctactttaag aaaattgaat gcttcgattc tgcgagatc	1740
tccgggtag aagatcgatt taatgcgtca cttggtagct atcatgacct cctaaagata	1800
attaaagata aggacttctt ggataacgaa gagaatgaag atatcttaga agatatagtg	1860
ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgt	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 gcacagggtt cgggacaagg ggactcattg	2160
cacgaacata ttgcgaatct tgctgggtcg ccagccatca aaaagggcat actccagaca	2220
gtcaaagtag tggatgagct agttaaggtc atgggacgac acaaacgga aaacattgta	2280
atcgagatgg cagcgaaaa tcaaacgact cagaaggggc aaaaaacag tcgagagcgg	2340
atgaagagaa tagaagaggg tattaagaa ctgggcagcc agatcttaa ggagcatcct	2400
gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg	2460
gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac	2520
attgtacccc aatcctttt gaaggacgat tcaatcgaca ataaagtgt tacacgctcg	2580
gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaa gaaaatgaag	2640
aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagt cgataactta	2700
actaaagctg agagggttg cttgtctgaa cttgacaagg cggatttat taaacgtcag	2760
ctcgtggaaa cccgccaat cacaagcat gttgcacaga tactagattc ccgaatgaat	2820
acgaaatcgc acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca	2880
aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac	2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa	3000
taccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag	3060
atgatcgca aaagcgaaca ggagataggc aaggctacag ccaataactt cttttattct	3120
aacattatga atttcttta gacggaaatc actctggcaa acggagagat acgcaaacga	3180
cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc	3240
gcgacggtga gaaaagtttt gtccatgcc caagtcaaca tagtaaagaa aactgaggtg	3300
cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc	3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc	3420

-continued

```

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 actacgtctt aaatacgtga atttcctgta tttagcgtcc 3720
cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact tttgttgag 3780
cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840
atcctagctg atgccaatct ggacaaagta ttaagcgcac acaacaagca cagggataaa 3900
cccatacgtg agcaggcgga aaatattatc catttggtta ctcttaccac 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> SEQ ID NO 4
<211> LENGTH: 1368
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

```

```

<400> SEQUENCE: 4

```

```

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

```

-continued

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		

-continued

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
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
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
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
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
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

-continued

Val 1010	Tyr	Gly	Asp	Tyr	Lys	Val 1015	Tyr	Asp	Val	Arg	Lys 1020	Met	Ile	Ala
Lys 1025	Ser	Glu	Gln	Glu	Ile	Gly 1030	Lys	Ala	Thr	Ala	Lys 1035	Tyr	Phe	Phe
Tyr 1040	Ser	Asn	Ile	Met	Asn	Phe 1045	Phe	Lys	Thr	Glu	Ile 1050	Thr	Leu	Ala
Asn 1055	Gly	Glu	Ile	Arg	Lys	Arg 1060	Pro	Leu	Ile	Glu	Thr 1065	Asn	Gly	Glu
Thr 1070	Gly	Glu	Ile	Val	Trp	Asp 1075	Lys	Gly	Arg	Asp	Phe 1080	Ala	Thr	Val
Arg 1085	Lys	Val	Leu	Ser	Met	Pro 1090	Gln	Val	Asn	Ile	Val 1095	Lys	Lys	Thr
Glu 1100	Val	Gln	Thr	Gly	Gly	Phe 1105	Ser	Lys	Glu	Ser	Ile 1110	Leu	Pro	Lys
Arg 1115	Asn	Ser	Asp	Lys	Leu	Ile 1120	Ala	Arg	Lys	Lys	Asp 1125	Trp	Asp	Pro
Lys 1130	Lys	Tyr	Gly	Gly	Phe	Asp 1135	Ser	Pro	Thr	Val	Ala 1140	Tyr	Ser	Val
Leu 1145	Val	Val	Ala	Lys	Val	Glu 1150	Lys	Gly	Lys	Ser	Lys 1155	Lys	Leu	Lys
Ser 1160	Val	Lys	Glu	Leu	Leu	Gly 1165	Ile	Thr	Ile	Met	Glu 1170	Arg	Ser	Ser
Phe 1175	Glu	Lys	Asn	Pro	Ile	Asp 1180	Phe	Leu	Glu	Ala	Lys 1185	Gly	Tyr	Lys
Glu 1190	Val	Lys	Lys	Asp	Leu	Ile 1195	Ile	Lys	Leu	Pro	Lys 1200	Tyr	Ser	Leu
Phe 1205	Glu	Leu	Glu	Asn	Gly	Arg 1210	Lys	Arg	Met	Leu	Ala 1215	Ser	Ala	Gly
Glu 1220	Leu	Gln	Lys	Gly	Asn	Glu 1225	Leu	Ala	Leu	Pro	Ser 1230	Lys	Tyr	Val
Asn 1235	Phe	Leu	Tyr	Leu	Ala	Ser 1240	His	Tyr	Glu	Lys	Leu 1245	Lys	Gly	Ser
Pro 1250	Glu	Asp	Asn	Glu	Gln	Lys 1255	Gln	Leu	Phe	Val	Glu 1260	Gln	His	Lys
His 1265	Tyr	Leu	Asp	Glu	Ile	Ile 1270	Glu	Gln	Ile	Ser	Glu 1275	Phe	Ser	Lys
Arg 1280	Val	Ile	Leu	Ala	Asp	Ala 1285	Asn	Leu	Asp	Lys	Val 1290	Leu	Ser	Ala
Tyr 1295	Asn	Lys	His	Arg	Asp	Lys 1300	Pro	Ile	Arg	Glu	Gln 1305	Ala	Glu	Asn
Ile 1310	Ile	His	Leu	Phe	Thr	Leu 1315	Thr	Asn	Leu	Gly	Ala 1320	Pro	Ala	Ala
Phe 1325	Lys	Tyr	Phe	Asp	Thr	Thr 1330	Ile	Asp	Arg	Lys	Arg 1335	Tyr	Thr	Ser
Thr 1340	Lys	Glu	Val	Leu	Asp	Ala 1345	Thr	Leu	Ile	His	Gln 1350	Ser	Ile	Thr
Gly 1355	Leu	Tyr	Glu	Thr	Arg	Ile 1360	Asp	Leu	Ser	Gln	Leu 1365	Gly	Gly	Asp

<210> SEQ ID NO 5

<211> LENGTH: 5

<212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 5

Glu Ala Ala Ala Lys
1 5

<210> SEQ ID NO 6

<211> LENGTH: 198

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 6

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> SEQ ID NO 7

<211> LENGTH: 198

<212> TYPE: PRT

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 7

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

-continued

```

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65          70          75
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> SEQ ID NO 8
<211> LENGTH: 198
<212> TYPE: PRT
<213> ORGANISM: Canis lupus

```

```

<400> SEQUENCE: 8

```

```

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> SEQ ID NO 9

```

-continued

<211> LENGTH: 199

<212> TYPE: PRT

<213> ORGANISM: Bos taurus

<400> SEQUENCE: 9

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> SEQ ID NO 10

<211> LENGTH: 429

<212> TYPE: PRT

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 10

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

-continued

```

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> SEQ ID NO 11

<211> LENGTH: 429

<212> TYPE: PRT

<213> ORGANISM: Rattus norvegicus

<400> SEQUENCE: 11

```

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 Arg Tyr Ala Ile Asp 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

```

-continued

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	Val
			100					105					110		
Leu	Arg	Phe	Leu	Ala	Thr	His	His	Asn	Leu	Ser	Leu	Asp	Ile	Phe	Ser
		115					120					125			
Ser	Arg	Leu	Tyr	Asn	Ile	Arg	Asp	Pro	Glu	Asn	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	Lys	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	Arg	Arg	Arg	Val	His	Leu	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	Gly	Val	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	Ile	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	His	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> SEQ ID NO 12

<211> LENGTH: 370

<212> TYPE: PRT

<213> ORGANISM: Macaca mulatta

-continued

<400> SEQUENCE: 12

```

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

```

-continued

```

<210> SEQ ID NO 13
<211> LENGTH: 384
<212> TYPE: PRT
<213> ORGANISM: Pan troglodytes

<400> SEQUENCE: 13

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

```

-continued

Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
370 375 380

<210> SEQ ID NO 14

<211> LENGTH: 377

<212> TYPE: PRT

<213> ORGANISM: Chlorocebus sabaeus

<400> SEQUENCE: 14

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

-continued

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> SEQ ID NO 15

<211> LENGTH: 384

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 15

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

<400> SEQUENCE: 16

Met 1	Lys	Pro	His	Phe 5	Arg	Asn	Thr	Val	Glu 10	Arg	Met	Tyr	Arg	Asp 15	Thr
Phe	Ser	Tyr	Asn 20	Phe	Tyr	Asn	Arg	Pro 25	Ile	Leu	Ser	Arg	Arg 30	Asn	Thr
Val	Trp	Leu 35	Cys	Tyr	Glu	Val	Lys 40	Thr	Lys	Gly	Pro	Ser 45	Arg	Pro	Arg
Leu	Asp 50	Ala	Lys	Ile	Phe	Arg 55	Gly	Gln	Val	Tyr	Ser 60	Gln	Pro	Glu	His
His 65	Ala	Glu	Met	Cys	Phe 70	Leu	Ser	Trp	Phe	Cys 75	Gly	Asn	Gln	Leu	Pro 80
Ala	Tyr	Lys	Cys 85	Phe	Gln	Ile	Thr	Trp 90	Phe	Val	Ser	Trp	Thr	Pro 95	Cys
Pro	Asp	Cys	Val 100	Ala	Lys	Leu	Ala	Glu 105	Phe	Leu	Ala	Glu	His 110	Pro	Asn
Val	Thr	Leu 115	Thr	Ile	Ser	Ala	Ala 120	Arg	Leu	Tyr	Tyr	Tyr 125	Trp	Glu	Arg
Asp 130	Tyr	Arg	Arg	Ala	Leu 135	Cys	Arg	Leu	Ser	Gln	Ala 140	Gly	Ala	Arg	Val
Lys 145	Ile	Met	Asp	Asp	Glu 150	Glu	Phe	Ala	Tyr	Cys 155	Trp	Glu	Asn	Phe	Val 160
Tyr	Ser	Glu	Gly	Gln 165	Pro	Phe	Met	Pro	Trp	Tyr	Lys	Phe	Asp	Asp 175	Asn
Tyr	Ala	Phe	Leu 180	His	Arg	Thr	Leu	Lys 185	Glu	Ile	Leu	Arg	Asn 190	Pro	Met
Glu	Ala	Met	Tyr 195	Pro	His	Ile	Phe	Tyr 200	Phe	His	Phe	Lys 205	Asn	Leu	Arg
Lys 210	Ala	Tyr	Gly	Arg	Asn 215	Glu	Ser	Trp	Leu	Cys	Phe	Thr	Met	Glu	Val
Val 225	Lys	His	His	Ser	Pro 230	Val	Ser	Trp	Lys	Arg	Gly	Val	Phe	Arg	Asn 240
Gln	Val	Asp	Pro	Glu 245	Thr	His	Cys	His	Ala	Glu	Arg	Cys	Phe	Leu 255	Ser
Trp	Phe	Cys	Asp 260	Asp	Ile	Leu	Ser	Pro 265	Asn	Thr	Asn	Tyr	Glu 270	Val	Thr
Trp	Tyr	Thr	Ser 275	Trp	Ser	Pro	Cys	Pro 280	Glu	Cys	Ala	Gly	Glu 285	Val	Ala
Glu 290	Phe	Leu	Ala	Arg	His	Ser	Asn 295	Val	Asn	Leu	Thr	Ile	Phe	Thr	Ala

-continued

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> SEQ ID NO 17
 <211> LENGTH: 382
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17

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			

-continued

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> SEQ ID NO 18
 <211> LENGTH: 190
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

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> SEQ ID NO 19
 <211> LENGTH: 199
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

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

-continued

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> SEQ ID NO 20

<211> LENGTH: 200

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 20

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

-continued

180	185	190
Asp Ile Leu Cys Asp Ala Glu Val		
195	200	
 <210> SEQ ID NO 21		
<211> LENGTH: 386		
<212> TYPE: PRT		
<213> ORGANISM: Homo sapiens		
 <400> SEQUENCE: 21		
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

```
<210> SEQ ID NO 22
<211> LENGTH: 236
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 22
```

```
<210> SEQ ID NO 23
<211> LENGTH: 229
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 23
```

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

-continued

```

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> SEQ ID NO 24

<211> LENGTH: 229

<212> TYPE: PRT

<213> ORGANISM: Rattus norvegicus

<400> SEQUENCE: 24

```

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

```

-continued

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> SEQ ID NO 25
 <211> LENGTH: 191
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 25

Met Glu Ala Lys Ala Ala Pro Lys Pro Ala Ala Ser Gly Ala Cys Ser
 1 5 10 15
 Val Ser Ala Glu Glu Thr Glu Lys Trp Met Glu Glu Ala Met His Met
 20 25 30
 Ala Lys Glu Ala Leu Glu Asn Thr Glu Val Pro Val Gly Cys Leu Met
 35 40 45
 Val Tyr Asn Asn Glu Val Val Gly Lys Gly Arg Asn Glu Val Asn Gln
 50 55 60
 Thr Lys Asn Ala Thr Arg His Ala Glu Met Val Ala Ile Asp Gln Val
 65 70 75 80
 Leu Asp Trp Cys Arg Gln Ser Gly Lys Ser Pro Ser Glu Val Phe Glu
 85 90 95
 His Thr Val Leu Tyr Val Thr Val Glu Pro Cys Ile Met Cys Ala Ala
 100 105 110
 Ala Leu Arg Leu Met Lys Ile Pro Leu Val Val Tyr Gly Cys Gln Asn
 115 120 125
 Glu Arg Phe Gly Gly Cys Gly Ser Val Leu Asn Ile Ala Ser Ala Asp
 130 135 140
 Leu Pro Asn Thr Gly Arg Pro Phe Gln Cys Ile Pro Gly Tyr Arg Ala
 145 150 155 160
 Glu Glu Ala Val Glu Met Leu Lys Thr Phe Tyr Lys Gln Glu Asn Pro
 165 170 175
 Asn Ala Pro Lys Ser Lys Val Arg Lys Lys Glu Cys Gln Lys Ser
 180 185 190

<210> SEQ ID NO 26
 <211> LENGTH: 191
 <212> TYPE: PRT
 <213> ORGANISM: Mus musculus

<400> SEQUENCE: 26

Met Glu Glu Lys Val Glu Ser Thr Thr Thr Pro Asp Gly Pro Cys Val
 1 5 10 15
 Val Ser Val Gln Glu Thr Glu Lys Trp Met Glu Glu Ala Met Arg Met
 20 25 30

-continued

Ala Lys Glu Ala Leu Glu Asn Ile Glu Val Pro Val Gly Cys Leu Met
 35 40 45

Val Tyr Asn Asn Glu Val Val Gly Lys Gly Arg Asn Glu Val Asn Gln
 50 55 60

Thr Lys Asn Ala Thr Arg His Ala Glu Met Val Ala Ile Asp Gln Val
 65 70 75 80

Leu Asp Trp Cys His Gln His Gly Gln Ser Pro Ser Thr Val Phe Glu
 85 90 95

His Thr Val Leu Tyr Val Thr Val Glu Pro Cys Ile Met Cys Ala Ala
 100 105 110

Ala Leu Arg Leu Met Lys Ile Pro Leu Val Val Tyr Gly Cys Gln Asn
 115 120 125

Glu Arg Phe Gly Gly Cys Gly Ser Val Leu Asn Ile Ala Ser Ala Asp
 130 135 140

Leu Pro Asn Thr Gly Arg Pro Phe Gln Cys Ile Pro Gly Tyr Arg Ala
 145 150 155 160

Glu Glu Ala Val Glu Leu Leu Lys Thr Phe Tyr Lys Gln Glu Asn Pro
 165 170 175

Asn Ala Pro Lys Ser Lys Val Arg Lys Lys Asp Cys Gln Lys Ser
 180 185 190

<210> SEQ ID NO 27

<211> LENGTH: 499

<212> TYPE: PRT

<213> ORGANISM: Mus musculus

<400> SEQUENCE: 27

Met Trp Thr Ala Asp Glu Ile Ala Gln Leu Cys Tyr Ala His Tyr Asn
 1 5 10 15

Val Arg Leu Pro Lys Gln Gly Lys Pro Glu Pro Asn Arg Glu Trp Thr
 20 25 30

Leu Leu Ala Ala Val Val Lys Ile Gln Ala Ser Ala Asn Gln Ala Cys
 35 40 45

Asp Ile Pro Glu Lys Glu Val Gln Val Thr Lys Glu Val Val Ser Met
 50 55 60

Gly Thr Gly Thr Lys Cys Ile Gly Gln Ser Lys Met Arg Glu Ser Gly
 65 70 75 80

Asp Ile Leu Asn Asp Ser His Ala Glu Ile Ile Ala Arg Arg Ser Phe
 85 90 95

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

Asp Ser Ile Phe Val Pro Gly Thr Gln Arg Gly Leu Trp Arg Leu Arg
 115 120 125

Pro Asp Leu Ser Phe Val Phe Phe Ser Ser His Thr Pro Cys Gly Asp
 130 135 140

Ala Ser Ile Ile Pro Met Leu Glu Phe Glu Glu Gln Pro Cys Cys Pro
 145 150 155 160

Val Ile Arg Ser Trp Ala Asn Asn Ser Pro Val Gln Glu Thr Glu Asn
 165 170 175

Leu Glu Asp Ser Lys Asp Lys Arg Asn Cys Glu Asp Pro Ala Ser Pro
 180 185 190

Val Ala Lys Lys Met Arg Leu Gly Thr Pro Ala Arg Ser Leu Ser Asn

-continued

195					200					205					
Cys	Val	Ala	His	His	Gly	Thr	Gln	Glu	Ser	Gly	Pro	Val	Lys	Pro	Asp
210					215						220				
Val	Ser	Ser	Ser	Asp	Leu	Thr	Lys	Glu	Glu	Pro	Asp	Ala	Ala	Asn	Gly
225					230					235					240
Ile	Ala	Ser	Gly	Ser	Phe	Arg	Val	Val	Asp	Val	Tyr	Arg	Thr	Gly	Ala
				245					250					255	
Lys	Cys	Val	Pro	Gly	Glu	Thr	Gly	Asp	Leu	Arg	Glu	Pro	Gly	Ala	Ala
			260					265					270		
Tyr	His	Gln	Val	Gly	Leu	Leu	Arg	Val	Lys	Pro	Gly	Arg	Gly	Asp	Arg
	275						280					285			
Thr	Cys	Ser	Met	Ser	Cys	Ser	Asp	Lys	Met	Ala	Arg	Trp	Asn	Val	Leu
	290						295				300				
Gly	Cys	Gln	Gly	Ala	Leu	Leu	Met	His	Phe	Leu	Glu	Lys	Pro	Ile	Tyr
305					310					315					320
Leu	Ser	Ala	Val	Val	Ile	Gly	Lys	Cys	Pro	Tyr	Ser	Gln	Glu	Ala	Met
				325					330					335	
Arg	Arg	Ala	Leu	Thr	Gly	Arg	Cys	Glu	Glu	Thr	Leu	Val	Leu	Pro	Arg
			340					345					350		
Gly	Phe	Gly	Val	Gln	Glu	Leu	Glu	Ile	Gln	Gln	Ser	Gly	Leu	Leu	Phe
	355					360						365			
Glu	Gln	Ser	Arg	Cys	Ala	Val	His	Arg	Lys	Arg	Gly	Asp	Ser	Pro	Gly
	370					375					380				
Arg	Leu	Val	Pro	Cys	Gly	Ala	Ala	Ile	Ser	Trp	Ser	Ala	Val	Pro	Gln
385					390					395					400
Gln	Pro	Leu	Asp	Val	Thr	Ala	Asn	Gly	Phe	Pro	Gln	Gly	Thr	Thr	Lys
				405					410					415	
Lys	Glu	Ile	Gly	Ser	Pro	Arg	Ala	Arg	Ser	Arg	Ile	Ser	Lys	Val	Glu
			420					425					430		
Leu	Phe	Arg	Ser	Phe	Gln	Lys	Leu	Leu	Ser	Ser	Ile	Ala	Asp	Asp	Glu
	435					440						445			
Gln	Pro	Asp	Ser	Ile	Arg	Val	Thr	Lys	Lys	Leu	Asp	Thr	Tyr	Gln	Glu
	450					455					460				
Tyr	Lys	Asp	Ala	Ala	Ser	Ala	Tyr	Gln	Glu	Ala	Trp	Gly	Ala	Leu	Arg
465					470					475					480
Arg	Ile	Gln	Pro	Phe	Ala	Ser	Trp	Ile	Arg	Asn	Pro	Pro	Asp	Tyr	His
				485					490					495	

Gln Phe Lys

<210> SEQ ID NO 28

<211> LENGTH: 502

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 28

Met	Trp	Thr	Ala	Asp	Glu	Ile	Ala	Gln	Leu	Cys	Tyr	Glu	His	Tyr	Gly
1				5					10					15	
Ile	Arg	Leu	Pro	Lys	Lys	Gly	Lys	Pro	Glu	Pro	Asn	His	Glu	Trp	Thr
		20						25					30		
Leu	Leu	Ala	Ala	Val	Val	Lys	Ile	Gln	Ser	Pro	Ala	Asp	Lys	Ala	Cys
		35					40					45			
Asp	Thr	Pro	Asp	Lys	Pro	Val	Gln	Val	Thr	Lys	Glu	Val	Val	Ser	Met

-continued

50					55					60					
Gly 65	Thr	Gly	Thr	Lys	Cys 70	Ile	Gly	Gln	Ser	Lys 75	Met	Arg	Lys	Asn	Gly 80
Asp	Ile	Leu	Asn 85	Asp	Ser	His	Ala	Glu	Val 90	Ile	Ala	Arg	Arg	Ser 95	Phe
Gln	Arg	Tyr	Leu 100	Leu	His	Gln	Leu	Gln 105	Leu	Ala	Ala	Thr	Leu 110	Lys	Glu
Asp	Ser	Ile 115	Phe	Val	Pro	Gly	Thr 120	Gln	Lys	Gly	Val	Trp 125	Lys	Leu	Arg
Arg	Asp 130	Leu	Ile	Phe	Val	Phe 135	Phe	Ser	Ser	His	Thr 140	Pro	Cys	Gly	Asp
Ala 145	Ser	Ile	Ile	Pro	Met 150	Leu	Glu	Phe	Glu	Asp 155	Gln	Pro	Cys	Cys	Pro 160
Val	Phe	Arg	Asn 165	Trp	Ala	His	Asn	Ser	Ser 170	Val	Glu	Ala	Ser	Ser 175	Asn
Leu	Glu	Ala	Pro 180	Gly	Asn	Glu	Arg	Lys 185	Cys	Glu	Asp	Pro	Asp 190	Ser	Pro
Val	Thr	Lys 195	Lys	Met	Arg	Leu	Glu 200	Pro	Gly	Thr	Ala	Ala 205	Arg	Glu	Val
Thr	Asn 210	Gly	Ala	Ala	His	His 215	Gln	Ser	Phe	Gly	Lys 220	Gln	Lys	Ser	Gly
Pro 225	Ile	Ser	Pro	Gly	Ile 230	His	Ser	Cys	Asp	Leu 235	Thr	Val	Glu	Gly	Leu 240
Ala	Thr	Val	Thr 245	Arg	Ile	Ala	Pro	Gly	Ser 250	Ala	Lys	Val	Ile	Asp 255	Val
Tyr	Arg	Thr	Gly 260	Ala	Lys	Cys	Val	Pro 265	Gly	Glu	Ala	Gly	Asp 270	Ser	Gly
Lys	Pro	Gly 275	Ala	Ala	Phe	His	Gln 280	Val	Gly	Leu	Leu	Arg 285	Val	Lys	Pro
Gly	Arg 290	Gly	Asp	Arg	Thr	Arg 295	Ser	Met	Ser	Cys	Ser 300	Asp	Lys	Met	Ala
Arg 305	Trp	Asn	Val	Leu	Gly 310	Cys	Gln	Gly	Ala	Leu 315	Leu	Met	His	Leu	Leu 320
Glu	Glu	Pro	Ile 325	Tyr	Leu	Ser	Ala	Val	Val 330	Ile	Gly	Lys	Cys	Pro 335	Tyr
Ser	Gln	Glu	Ala 340	Met	Gln	Arg	Ala	Leu 345	Ile	Gly	Arg	Cys	Gln 350	Asn	Val
Ser	Ala	Leu 355	Pro	Lys	Gly	Phe	Gly 360	Val	Gln	Glu	Leu	Lys 365	Ile	Leu	Gln
Ser	Asp 370	Leu	Leu	Phe	Glu	Gln 375	Ser	Arg	Ser	Ala	Val 380	Gln	Ala	Lys	Arg
Ala 385	Asp	Ser	Pro	Gly	Arg 390	Leu	Val	Pro	Cys	Gly 395	Ala	Ala	Ile	Ser	Trp 400
Ser	Ala	Val	Pro 405	Glu	Gln	Pro	Leu	Asp	Val 410	Thr	Ala	Asn	Gly	Phe	Pro 415
Gln	Gly	Thr	Thr 420	Lys	Lys	Thr	Ile	Gly 425	Ser	Leu	Gln	Ala	Arg	Ser	Gln
Ile	Ser	Lys 435	Val	Glu	Leu	Phe	Arg 440	Ser	Phe	Gln	Lys	Leu 445	Leu	Ser	Arg
Ile	Ala 450	Arg	Asp	Lys	Trp 455	Pro	His	Ser	Leu	Arg	Val 460	Gln	Lys	Leu	Asp

-continued

Thr Tyr Gln Glu Tyr Lys Glu Ala Ala Ser Ser Tyr Gln Glu Ala Trp
465 470 475 480

Ser Thr Leu Arg Lys Gln Val Phe Gly Ser Trp Ile Arg Asn Pro Pro
485 490 495

Asp Tyr His Gln Phe Lys
500

<210> SEQ ID NO 29
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 29

Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser
1 5 10

<210> SEQ ID NO 30
<211> LENGTH: 1580
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 30

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 Asp Lys
20 25 30

Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala
35 40 45

Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu
50 55 60

Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu
65 70 75 80

Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr
85 90 95

Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln
100 105 110

Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His
115 120 125

Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu Arg
130 135 140

His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu Lys
145 150 155 160

Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr Asp
165 170 175

Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys
180 185 190

Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn Ser
195 200 205

Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln Leu
210 215 220

Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala Ile

-continued

225					230						235				240
Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala
				245					250					255	
Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala
			260					265					270		
Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala
		275					280					285			
Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu
	290					295					300				
Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu
305					310					315					320
Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg
			325						330					335	
Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys
			340					345					350		
Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val
	355						360					365			
Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser
	370					375					380				
Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu
385					390					395					400
Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu
			405						410					415	
Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg
		420						425					430		
Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu
		435					440					445			
His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp
	450					455					460				
Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr
465					470					475					480
Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg
			485						490					495	
Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp
		500						505					510		
Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp
		515					520					525			
Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr
	530					535					540				
Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr
545					550					555					560
Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala
			565						570					575	
Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln
		580						585					590		
Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu
		595					600					605			
Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His
	610					615					620				
Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu
625					630					635					640

-continued

Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	
				645					650					655		
Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	
			660					665					670			
Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	
		675					680					685				
Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	
	690				695						700					
Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	
705				710						715				720		
Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	
			725					730						735		
Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	
	740						745						750			
Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	
	755					760						765				
Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys	
	770				775					780						
Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln	
785				790						795				800		
Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu	Glu	Gly	
			805					810					815			
Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	Val	Glu	Asn	
	820						825					830				
Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	
	835					840					845					
Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	Leu	Ser	Asp	
	850				855					860						
Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	Asp	Asp	Ser	
865				870						875				880		
Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	Gly	Lys	Ser	
			885					890					895			
Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp	
		900					905					910				
Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	Phe	Asp	Asn	
	915					920					925					
Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	Lys	Ala	Gly	
	930				935					940						
Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	
945				950					955					960		
Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	Glu	Asn	Asp	
			965				970						975			
Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	Lys	Leu	Val	
		980				985						990				
Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	Glu	Ile	Asn	
	995					1000					1005					
Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Val	Gly		
1010					1015						1020					
Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	Val		
1025					1030						1035					

-continued

Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala	Lys
1040						1045					1050			
Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe	Tyr
1055						1060					1065			
Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala	Asn
1070						1075					1080			
Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu	Thr
1085						1090					1095			
Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val	Arg
1100						1105					1110			
Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr	Glu
1115						1120					1125			
Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	Arg
1130						1135					1140			
Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro	Lys
1145						1150					1155			
Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val	Leu
1160						1165					1170			
Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	Ser
1175						1180					1185			
Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser	Phe
1190						1195					1200			
Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys	Glu
1205						1210					1215			
Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe
1220						1225					1230			
Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly	Glu
1235						1240					1245			
Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val	Asn
1250						1255					1260			
Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser	Pro
1265						1270					1275			
Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys	His
1280						1285					1290			
Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys	Arg
1295						1300					1305			
Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala	Tyr
1310						1315					1320			
Asn	Lys	His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu	Asn	Ile
1325						1330					1335			
Ile	His	Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala	Phe
1340						1345					1350			
Lys	Tyr	Phe	Asp	Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr	Ser	Thr
1355						1360					1365			
Lys	Glu	Val	Leu	Asp	Ala	Thr	Leu	Ile	His	Gln	Ser	Ile	Thr	Gly
1370						1375					1380			
Leu	Tyr	Glu	Thr	Arg	Ile	Asp	Leu	Ser	Gln	Leu	Gly	Gly	Asp	Gly
1385						1390					1395			
Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser	Tyr
1400						1405					1410			
Val	Val	Lys	Arg	Arg	Asp	Ser	Ala	Thr	Ser	Phe	Ser	Leu	Asp	Phe

-continued

1415	1420	1425
Gly Tyr Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe		
1430	1435	1440
Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr		
1445	1450	1455
Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala		
1460	1465	1470
Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu		
1475	1480	1485
Arg Ile Phe Thr Ala Arg Leu Tyr Phe Cys Glu Asp Arg Lys Ala		
1490	1495	1500
Glu Pro Glu Gly Leu Arg Arg Leu His Arg Ala Gly Val Gln Ile		
1505	1510	1515
Ala Ile Met Thr Phe Lys Asp Tyr Phe Tyr Cys Trp Asn Thr Phe		
1520	1525	1530
Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His		
1535	1540	1545
Glu Asn Ser Val Arg Leu Ser Arg Gln Leu Arg Arg Ile Leu Leu		
1550	1555	1560
Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu		
1565	1570	1575
Gly Leu		
1580		

<210> SEQ ID NO 31
 <211> LENGTH: 1564
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 31

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys		
1	5	10
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
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
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu		

-continued

165					170					175				
Arg	Arg	Ile	Leu	Leu	Pro	Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Ser
			180					185				190		
Gly	Gly	Gly	Gly	Ser	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile
		195					200					205		Gly
Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val
	210					215				220				Pro
Ser	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile
225				230						235				240
Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala
			245						250					255
Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Lys
		260						265					270	
Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala
	275						280					285		Lys
Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val
	290					295					300			Glu
Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val
305				310						315				320
Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg
			325						330					335
Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr
		340						345					350	Leu
Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu
		355					360					365		Gly
Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln
	370					375					380			Leu
Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala
385				390						395				Ser
Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser
			405						410					Arg
Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn
		420						425					430	Gly
Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn
	435						440					445		Phe
Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser
	450					455					460			Lys
Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly
465				470						475				480
Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala
			485						490					495
Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala
		500						505					510	Pro
Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp
		515					520					525		Leu
Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr
	530					535						540		Lys
Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile
545					550					555				560
Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile
			565						570					575

Glu 580	Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu 585	Leu	Val	Lys	Leu 590	Asn	Arg	Glu
Asp 595	Leu	Leu	Arg	Lys	Gln	Arg	Thr 600	Phe	Asp	Asn	Gly	Ser 605	Ile	Pro	His
Gln 610	Ile	His	Leu	Gly	Glu	Leu 615	His	Ala	Ile	Leu	Arg 620	Arg	Gln	Glu	Asp
Phe 625	Tyr	Pro	Phe	Leu	Lys 630	Asp	Asn	Arg	Glu	Lys 635	Ile	Glu	Lys	Ile	Leu 640
Thr	Phe	Arg	Ile	Pro 645	Tyr	Tyr	Val	Gly	Pro 650	Leu	Ala	Arg	Gly	Asn 655	Ser
Arg	Phe	Ala	Trp 660	Met	Thr	Arg	Lys	Ser 665	Glu	Glu	Thr	Ile	Thr 670	Pro	Trp
Asn	Phe	Glu	Glu	Val 675	Val	Asp	Lys 680	Gly	Ala	Ser	Ala	Gln 685	Ser	Phe	Ile
Glu 690	Arg	Met	Thr	Asn	Phe	Asp 695	Lys	Asn	Leu	Pro	Asn 700	Glu	Lys	Val	Leu
Pro 705	Lys	His	Ser	Leu	Leu 710	Tyr	Glu	Tyr	Phe	Thr 715	Val	Tyr	Asn	Glu	Leu 720
Thr	Lys	Val	Lys	Tyr 725	Val	Thr	Glu	Gly	Met 730	Arg	Lys	Pro	Ala	Phe 735	Leu
Ser	Gly	Glu	Gln 740	Lys	Lys	Ala	Ile	Val 745	Asp	Leu	Leu	Phe 750	Lys	Thr	Asn
Arg	Lys	Val 755	Thr	Val	Lys	Gln	Leu 760	Lys	Glu	Asp	Tyr	Phe 765	Lys	Lys	Ile
Glu 770	Cys	Phe	Asp	Ser	Val 775	Glu	Ile	Ser	Gly	Val 780	Glu	Asp	Arg	Phe	Asn
Ala 785	Ser	Leu	Gly	Thr	Tyr 790	His	Asp	Leu	Leu	Lys 795	Ile	Ile	Lys	Asp	Lys 800
Asp	Phe	Leu	Asp 805	Asn	Glu	Glu	Asn	Glu	Asp 810	Ile	Leu	Glu	Asp	Ile 815	Val
Leu	Thr	Leu	Thr 820	Leu	Phe	Glu	Asp	Arg 825	Glu	Met	Ile	Glu	Glu 830	Arg	Leu
Lys	Thr	Tyr 835	Ala	His	Leu	Phe	Asp 840	Asp	Lys	Val	Met	Lys 845	Gln	Leu	Lys
Arg	Arg 850	Arg	Tyr	Thr	Gly	Trp 855	Gly	Arg	Leu	Ser	Arg 860	Lys	Leu	Ile	Asn
Gly 865	Ile	Arg	Asp	Lys	Gln 870	Ser	Gly	Lys	Thr	Ile 875	Leu	Asp	Phe	Leu	Lys 880
Ser	Asp	Gly	Phe 885	Ala	Asn	Arg	Asn	Phe	Met 890	Gln	Leu	Ile	His	Asp 895	Asp
Ser	Leu	Thr	Phe 900	Lys	Glu	Asp	Ile	Gln 905	Lys	Ala	Gln	Val 910	Ser	Gly	Gln
Gly	Asp 915	Ser	Leu	His	Glu	His	Ile 920	Ala	Asn	Leu	Ala	Gly 925	Ser	Pro	Ala
Ile 930	Lys	Lys	Gly	Ile	Leu	Gln 935	Thr	Val	Lys	Val 940	Val	Asp	Glu	Leu	Val
Lys 945	Val	Met	Gly	Arg	His 950	Lys	Pro	Glu	Asn	Ile 955	Val	Ile	Glu	Met	Ala 960
Arg	Glu	Asn	Gln 965	Thr	Thr	Gln	Lys	Gly	Gln 970	Lys	Asn	Ser	Arg	Glu 975	Arg

-continued

Met	Lys	Arg	Ile	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	980	985	990	
Lys	Glu	His	Pro	Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	995	1000	1005
Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	1010	1015	1020	
Leu	Asp	Ile	Asn	Arg	Leu	Ser	Asp	Tyr	Asp	Val	Asp	Ala	Ile	Val	1025	1030	1035	
Pro	Gln	Ser	Phe	Leu	Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	1040	1045	1050	
Thr	Arg	Ser	Asp	Lys	Asn	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	1055	1060	1065	
Glu	Glu	Val	Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	1070	1075	1080	
Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	Phe	Asp	Asn	Leu	Thr	Lys	1085	1090	1095	
Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	Lys	Ala	Gly	Phe	Ile	1100	1105	1110	
Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	Ala	1115	1120	1125	
Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	Glu	Asn	Asp	1130	1135	1140	
Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	Lys	Leu	1145	1150	1155	
Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	Glu	1160	1165	1170	
Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	1175	1180	1185	
Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	1190	1195	1200	
Phe	Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	1205	1210	1215	
Ala	Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	1220	1225	1230	
Phe	Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	1235	1240	1245	
Ala	Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	1250	1255	1260	
Glu	Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	1265	1270	1275	
Val	Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	1280	1285	1290	
Thr	Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	1295	1300	1305	
Lys	Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	1310	1315	1320	
Pro	Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	1325	1330	1335	
Val	Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	1340	1345	1350	
Lys	Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser				

-continued

1355	1360	1365
Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr		
1370	1375	1380
Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser		
1385	1390	1395
Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala		
1400	1405	1410
Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr		
1415	1420	1425
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly		
1430	1435	1440
Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His		
1445	1450	1455
Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser		
1460	1465	1470
Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser		
1475	1480	1485
Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu		
1490	1495	1500
Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala		
1505	1510	1515
Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr		
1520	1525	1530
Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile		
1535	1540	1545
Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly		
1550	1555	1560

Asp

<210> SEQ ID NO 32

<211> LENGTH: 1579

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 32

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 Asp Lys			
20	25	30	
Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala			
35	40	45	
Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu			
50	55	60	
Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu			
65	70	75	80
Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr			
85	90	95	
Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln			
100	105	110	
Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe His			
115	120	125	

-continued

Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg
130						135					140				
His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys
145					150					155					160
Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp
				165					170					175	
Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys
			180					185					190		
Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser
		195					200					205			
Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu
	210					215					220				
Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile
225					230					235					240
Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala
				245					250					255	
Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala
			260					265					270		
Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala
		275					280					285			
Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu
	290					295				300					
Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu
305					310					315					320
Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg
				325					330					335	
Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys
			340					345						350	
Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val
		355					360					365			
Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser
	370					375					380				
Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu
385					390					395					400
Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu
				405					410					415	
Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg
			420					425					430		
Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu
		435					440					445			
His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp
	450					455					460				
Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr
465					470					475					480
Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg
				485					490					495	
Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp
			500					505					510		
Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp
		515					520						525		

-continued

Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr
530						535				540					
Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr
545					550					555					560
Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala
				565					570					575	
Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln
			580					585					590		
Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu
		595					600					605			
Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His
610						615					620				
Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu
625					630					635					640
Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu
				645					650					655	
Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe
			660					665					670		
Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp
		675					680					685			
Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser
690						695					700				
Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg
705					710					715					720
Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp
				725					730					735	
Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His
			740					745					750		
Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln
		755					760					765			
Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys
						775					780				
Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln
785					790					795					800
Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu	Glu	Gly
				805					810					815	
Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	Val	Glu	Asn
			820					825					830		
Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly
		835					840					845			
Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	Leu	Ser	Asp
		850				855					860				
Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	Asp	Asp	Ser
865					870					875					880
Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	Gly	Lys	Ser
				885					890					895	
Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp
			900					905					910		
Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	Phe	Asp	Asn
		915					920					925			
Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	Lys	Ala	Gly

-continued

930	935	940
Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val 945	950	955 960
Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp 965	970	975
Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val 980	985	990
Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn 995	1000	1005
Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly 1010	1015	1020
Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val 1025	1030	1035
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys 1040	1045	1050
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr 1055	1060	1065
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn 1070	1075	1080
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr 1085	1090	1095
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg 1100	1105	1110
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu 1115	1120	1125
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg 1130	1135	1140
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys 1145	1150	1155
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu 1160	1165	1170
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser 1175	1180	1185
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe 1190	1195	1200
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu 1205	1210	1215
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe 1220	1225	1230
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu 1235	1240	1245
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn 1250	1255	1260
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro 1265	1270	1275
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His 1280	1285	1290
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg 1295	1300	1305
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr 1310	1315	1320

-continued

```

Asn Lys  His Arg Asp Lys Pro  Ile Arg Glu Gln Ala  Glu Asn Ile
1325                      1330                      1335

Ile His   Leu Phe Thr Leu Thr  Asn Leu Gly Ala Pro  Ala Ala Phe
1340                      1345                      1350

Lys Tyr   Phe Asp Thr Thr  Ile  Asp Arg Lys Arg Tyr  Thr Ser Thr
1355                      1360                      1365

Lys Glu   Val Leu Asp Ala Thr  Leu Ile His Gln Ser  Ile Thr Gly
1370                      1375                      1380

Leu Tyr   Glu Thr Arg Ile Asp  Leu Ser Gln Leu Gly  Gly Asp Gly
1385                      1390                      1395

Gly Gly   Gly Ser Gly Gly Gly  Gly Ser Gly Gly Gly  Gly Ser Tyr
1400                      1405                      1410

Val Val   Lys Arg Arg Asp Ser  Ala Thr Ser Cys Ser  Leu Asp Phe
1415                      1420                      1425

Gly His   Leu Arg Asn Lys Ser  Gly Cys His Val Glu  Leu Leu Phe
1430                      1435                      1440

Leu Arg   Tyr Ile Ser Asp Trp  Asp Leu Asp Pro Gly  Arg Cys Tyr
1445                      1450                      1455

Arg Val   Thr Trp Phe Thr Ser  Trp Ser Pro Cys Tyr  Asp Cys Ala
1460                      1465                      1470

Arg His   Val Ala Glu Phe Leu  Arg Trp Asn Pro Asn  Leu Ser Leu
1475                      1480                      1485

Arg Ile   Phe Thr Ala Arg Leu  Tyr Phe Cys Glu Asp  Arg Lys Ala
1490                      1495                      1500

Glu Pro   Glu Gly Leu Arg Arg  Leu His Arg Ala Gly  Val Gln Ile
1505                      1510                      1515

Gly Ile   Met Thr Phe Lys Asp  Tyr Phe Tyr Cys Trp  Asn Thr Phe
1520                      1525                      1530

Val Glu   Asn Arg Glu Arg Thr  Phe Lys Ala Trp Glu  Gly Leu His
1535                      1540                      1545

Glu Asn   Ser Val Arg Leu Thr  Arg Gln Leu Arg Arg  Ile Leu Leu
1550                      1555                      1560

Pro Leu   Tyr Glu Val Asp Asp  Leu Arg Asp Ala Phe  Arg Met Leu
1565                      1570                      1575

```

Gly

```

<210> SEQ ID NO 33
<211> LENGTH: 1724
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

```

<400> SEQUENCE: 33

```

Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Met Glu Leu Lys Tyr
1          5          10          15

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
          20          25          30

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
          35          40          45

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro
          50          55          60

```

-continued

Lys	Val	Thr	Leu	Thr	Ile	Phe	Val	Ala	Arg	Leu	Tyr	Tyr	Phe	Trp	Asp	65	70	75	80
Pro	Asp	Tyr	Gln	Glu	Ala	Leu	Arg	Ser	Leu	Cys	Gln	Lys	Arg	Asp	Gly	85	90	95	
Pro	Arg	Ala	Thr	Met	Lys	Ile	Met	Asn	Tyr	Asp	Glu	Phe	Gln	His	Cys	100	105	110	
Trp	Ser	Lys	Phe	Val	Tyr	Ser	Gln	Arg	Glu	Leu	Phe	Glu	Pro	Trp	Asn	115	120	125	
Asn	Leu	Pro	Lys	Tyr	Tyr	Ile	Leu	Leu	His	Ile	Met	Leu	Gly	Glu	Ile	130	135	140	
Leu	Arg	His	Ser	Met	Asp	Pro	Pro	Thr	Phe	Thr	Phe	Asn	Phe	Asn	Asn	145	150	155	160
Glu	Pro	Trp	Val	Arg	Gly	Arg	His	Glu	Thr	Tyr	Leu	Cys	Tyr	Glu	Val	165	170	175	
Glu	Arg	Met	His	Asn	Asp	Thr	Trp	Val	Leu	Leu	Asn	Gln	Arg	Arg	Gly	180	185	190	
Phe	Leu	Cys	Asn	Gln	Ala	Pro	His	Lys	His	Gly	Phe	Leu	Glu	Gly	Arg	195	200	205	
His	Ala	Glu	Leu	Cys	Phe	Leu	Asp	Val	Ile	Pro	Phe	Trp	Lys	Leu	Asp	210	215	220	
Leu	Asp	Gln	Asp	Tyr	Arg	Val	Thr	Cys	Phe	Thr	Ser	Trp	Ser	Pro	Cys	225	230	235	240
Phe	Ser	Cys	Ala	Gln	Glu	Met	Ala	Lys	Phe	Ile	Ser	Lys	Asn	Lys	His	245	250	255	
Val	Ser	Leu	Cys	Ile	Phe	Thr	Ala	Arg	Ile	Tyr	Asp	Asp	Gln	Gly	Arg	260	265	270	
Cys	Gln	Glu	Gly	Leu	Arg	Thr	Leu	Ala	Glu	Ala	Gly	Ala	Lys	Ile	Ser	275	280	285	
Ile	Met	Thr	Tyr	Ser	Glu	Phe	Lys	His	Cys	Trp	Asp	Thr	Phe	Val	Asp	290	295	300	
His	Gln	Gly	Cys	Pro	Phe	Gln	Pro	Trp	Asp	Gly	Leu	Asp	Glu	His	Ser	305	310	315	320
Gln	Asp	Leu	Ser	Gly	Arg	Leu	Arg	Ala	Ile	Leu	Gln	Asn	Gln	Glu	Asn	325	330	335	
Ser	Pro	Lys	Lys	Lys	Arg	Lys	Val	Glu	Ala	Ser	Ser	Pro	Lys	Lys	Lys	340	345	350	
Arg	Lys	Val	Glu	Ala	Ser	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile	Gly	355	360	365	
Thr	Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	370	375	380	
Ser	Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	385	390	395	400
Lys	Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	405	410	415	
Ala	Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	420	425	430	
Asn	Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	435	440	445	
Val	Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	450	455	460	
Glu	Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp				

-continued

465					470						475				480
Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys
				485					490					495	
Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu
			500					505					510		
Ala	Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly
		515					520					525			
Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu
	530					535					540				
Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser
545					550					555					560
Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg
			565						570					575	
Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly
		580						585					590		
Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe
	595						600					605			
Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys
	610					615					620				
Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp
625					630					635					640
Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile
			645						650					655	
Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro
		660					665						670		
Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu
	675						680					685			
Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys
	690					695					700				
Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp
705					710					715					720
Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu
			725						730					735	
Glu	Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu
		740						745					750		
Asp	Leu	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His
	755						760					765			
Gln	Ile	His	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Gln	Glu	Asp	
	770					775					780				
Phe	Tyr	Pro	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu
785					790					795					800
Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser
			805						810					815	
Arg	Phe	Ala	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp
		820						825					830		
Asn	Phe	Glu	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile
		835					840					845			
Glu	Arg	Met	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu
	850					855					860				
Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu
865					870					875					880

Thr	Lys	Val	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu
885															
Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn
900															
Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile
915															
Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn
930															
Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys
945															
Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val
965															
Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu
980															
Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys
995															
Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	
1010															
Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	
1025															
Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	
1040															
His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	
1055															
Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	Ile	Ala	Asn	Leu	
1070															
Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Thr	Val	Lys	
1085															
Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys	Pro	Glu	
1100															
Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln	Lys	
1115															
Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu	Glu	Gly	
1130															
Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	Val	Glu	
1145															
Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	
1160															
Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	
1175															
Leu	Ser	Asp	Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser	Phe	Leu	
1190															
Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	
1205															
Asn	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	
1220															
Lys	Met	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	
1235															
Thr	Gln	Arg	Lys	Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	
1250															

-continued

Leu Ser 1265	Glu Leu Asp Lys 1270	Ala Gly Phe Ile Lys Arg 1275	Gln Leu Val
Glu Thr 1280	Arg Gln Ile Thr Lys 1285	His Val Ala Gln Ile 1290	Leu Asp Ser
Arg Met 1295	Asn Thr Lys Tyr Asp 1300	Glu Asn Asp Lys Leu 1305	Ile Arg Glu
Val Lys 1310	Val Ile Thr Leu Lys 1315	Ser Lys Leu Val Ser 1320	Asp Phe Arg
Lys Asp 1325	Phe Gln Phe Tyr Lys 1330	Val Arg Glu Ile Asn 1335	Asn Tyr His
His Ala 1340	His Asp Ala Tyr Leu 1345	Asn Ala Val Val Gly 1350	Thr Ala Leu
Ile Lys 1355	Lys Tyr Pro Lys Leu 1360	Glu Ser Glu Phe Val 1365	Tyr Gly Asp
Tyr Lys 1370	Val Tyr Asp Val Arg 1375	Lys Met Ile Ala Lys 1380	Ser Glu Gln
Glu Ile 1385	Gly Lys Ala Thr Ala 1390	Lys Tyr Phe Phe Tyr 1395	Ser Asn Ile
Met Asn 1400	Phe Phe Lys Thr Glu 1405	Ile Thr Leu Ala Asn 1410	Gly Glu Ile
Arg Lys 1415	Arg Pro Leu Ile Glu 1420	Thr Asn Gly Glu Thr 1425	Gly Glu Ile
Val Trp 1430	Asp Lys Gly Arg Asp 1435	Phe Ala Thr Val Arg 1440	Lys Val Leu
Ser Met 1445	Pro Gln Val Asn Ile 1450	Val Lys Lys Thr Glu 1455	Val Gln Thr
Gly Gly 1460	Phe Ser Lys Glu Ser 1465	Ile Leu Pro Lys Arg 1470	Asn Ser Asp
Lys Leu 1475	Ile Ala Arg Lys Lys 1480	Asp Trp Asp Pro Lys 1485	Lys Tyr Gly
Gly Phe 1490	Asp Ser Pro Thr Val 1495	Ala Tyr Ser Val Leu 1500	Val Val Ala
Lys Val 1505	Glu Lys Gly Lys Ser 1510	Lys Lys Leu Lys Ser 1515	Val Lys Glu
Leu Leu 1520	Gly Ile Thr Ile Met 1525	Glu Arg Ser Ser Phe 1530	Glu Lys Asn
Pro Ile 1535	Asp Phe Leu Glu Ala 1540	Lys Gly Tyr Lys Glu 1545	Val Lys Lys
Asp Leu 1550	Ile Ile Lys Leu Pro 1555	Lys Tyr Ser Leu Phe 1560	Glu Leu Glu
Asn Gly 1565	Arg Lys Arg Met Leu 1570	Ala Ser Ala Gly Glu 1575	Leu Gln Lys
Gly Asn 1580	Glu Leu Ala Leu Pro 1585	Ser Lys Tyr Val Asn 1590	Phe Leu Tyr
Leu Ala 1595	Ser His Tyr Glu Lys 1600	Leu Lys Gly Ser Pro 1605	Glu Asp Asn
Glu Gln 1610	Lys Gln Leu Phe Val 1615	Glu Gln His Lys His 1620	Tyr Leu Asp
Glu Ile 1625	Ile Glu Gln Ile Ser 1630	Glu Phe Ser Lys Arg 1635	Val Ile Leu
Ala Asp	Ala Asn Leu Asp Lys	Val Leu Ser Ala Tyr	Asn Lys His

-continued

1640	1645	1650
Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu		
1655	1660	1665
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe		
1670	1675	1680
Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val		
1685	1690	1695
Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu		
1700	1705	1710
Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp		
1715	1720	

<210> SEQ ID NO 34
 <211> LENGTH: 1368
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 34

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		

-continued

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			

-continued

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
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
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
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
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

-continued

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> SEQ ID NO 35

<211> LENGTH: 1851

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 35

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	Ser	Met
			20				25					30			

-continued

Gly	Thr	Gly	Thr	Lys	Cys	Ile	Gly	Gln	Ser	Lys	Met	Arg	Lys	Asn	Gly
		35					40					45			
Asp	Ile	Leu	Asn	Asp	Ser	His	Ala	Glu	Val	Ile	Ala	Arg	Arg	Ser	Phe
	50					55					60				
Gln	Arg	Tyr	Leu	Leu	His	Gln	Leu	Gln	Leu	Ala	Ala	Thr	Leu	Lys	Glu
65					70					75				80	
Asp	Ser	Ile	Phe	Val	Pro	Gly	Thr	Gln	Lys	Gly	Val	Trp	Lys	Leu	Arg
				85					90					95	
Arg	Asp	Leu	Ile	Phe	Val	Phe	Phe	Ser	Ser	His	Thr	Pro	Cys	Gly	Asp
		100						105					110		
Ala	Ser	Ile	Ile	Pro	Met	Leu	Glu	Phe	Glu	Asp	Gln	Pro	Cys	Cys	Pro
		115					120					125			
Val	Phe	Arg	Asn	Trp	Ala	His	Asn	Ser	Ser	Val	Glu	Ala	Ser	Ser	Asn
	130					135					140				
Leu	Glu	Ala	Pro	Gly	Asn	Glu	Arg	Lys	Cys	Glu	Asp	Pro	Asp	Ser	Pro
145					150					155					160
Val	Thr	Lys	Lys	Met	Arg	Leu	Glu	Pro	Gly	Thr	Ala	Ala	Arg	Glu	Val
				165					170					175	
Thr	Asn	Gly	Ala	Ala	His	His	Gln	Ser	Phe	Gly	Lys	Gln	Lys	Ser	Gly
		180						185					190		
Pro	Ile	Ser	Pro	Gly	Ile	His	Ser	Cys	Asp	Leu	Thr	Val	Glu	Gly	Leu
	195					200						205			
Ala	Thr	Val	Thr	Arg	Ile	Ala	Pro	Gly	Ser	Ala	Lys	Val	Ile	Asp	Val
	210					215					220				
Tyr	Arg	Thr	Gly	Ala	Lys	Cys	Val	Pro	Gly	Glu	Ala	Gly	Asp	Ser	Gly
225					230					235					240
Lys	Pro	Gly	Ala	Ala	Phe	His	Gln	Val	Gly	Leu	Leu	Arg	Val	Lys	Pro
				245					250					255	
Gly	Arg	Gly	Asp	Arg	Thr	Arg	Ser	Met	Ser	Cys	Ser	Asp	Lys	Met	Ala
		260						265					270		
Arg	Trp	Asn	Val	Leu	Gly	Cys	Gln	Gly	Ala	Leu	Leu	Met	His	Leu	Leu
		275					280					285			
Glu	Glu	Pro	Ile	Tyr	Leu	Ser	Ala	Val	Val	Ile	Gly	Lys	Cys	Pro	Tyr
	290					295					300				
Ser	Gln	Glu	Ala	Met	Gln	Arg	Ala	Leu	Ile	Gly	Arg	Cys	Gln	Asn	Val
305					310					315					320
Ser	Ala	Leu	Pro	Lys	Gly	Phe	Gly	Val	Gln	Glu	Leu	Lys	Ile	Leu	Gln
				325					330					335	
Ser	Asp	Leu	Leu	Phe	Glu	Gln	Ser	Arg	Ser	Ala	Val	Gln	Ala	Lys	Arg
		340						345					350		
Ala	Asp	Ser	Pro	Gly	Arg	Leu	Val	Pro	Cys	Gly	Ala	Ala	Ile	Ser	Trp
	355						360					365			
Ser	Ala	Val	Pro	Glu	Gln	Pro	Leu	Asp	Val	Thr	Ala	Asn	Gly	Phe	Pro
	370					375					380				
Gln	Gly	Thr	Thr	Lys	Lys	Thr	Ile	Gly	Ser	Leu	Gln	Ala	Arg	Ser	Gln
385					390					395					400
Ile	Ser	Lys	Val	Glu	Leu	Phe	Arg	Ser	Phe	Gln	Lys	Leu	Leu	Ser	Arg
			405						410					415	
Ile	Ala	Arg	Asp	Lys	Trp	Pro	His	Ser	Leu	Arg	Val	Gln	Lys	Leu	Asp
		420						425					430		
Thr	Tyr	Gln	Glu	Tyr	Lys	Glu	Ala	Ala	Ser	Ser	Tyr	Gln	Glu	Ala	Trp

-continued

435					440					445					
Ser	Thr	Leu	Arg	Lys	Gln	Val	Phe	Gly	Ser	Trp	Ile	Arg	Asn	Pro	Pro
450						455					460				
Asp	Tyr	His	Gln	Phe	Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser	Gly
465					470					475					480
Gly	Gly	Gly	Ser	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile	Gly	Thr
				485					490					495	
Asn	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	Ser
			500					505					510		
Lys	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys
		515					520					525			
Asn	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala
530					535						540				
Thr	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn
545					550					555					560
Arg	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val
				565					570					575	
Asp	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu
			580					585					590		
Asp	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu
		595					600					605			
Val	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys
610						615					620				
Leu	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala
625					630				635						640
Leu	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp
				645				650						655	
Leu	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val
		660						665					670		
Gln	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly
		675				680						685			
Val	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg
	690					695					700				
Leu	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu
705					710				715						720
Phe	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys
				725				730						735	
Ser	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp
		740						745					750		
Thr	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln
		755				760						765			
Tyr	Ala	Asp	Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu
	770					775					780				
Leu	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu
785					790					795					800
Ser	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr
				805					810					815	
Leu	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu
			820					825					830		
Ile	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly
		835					840					845			

-continued

Gly	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu
850						855					860				
Lys	Met	Asp	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp
865					870					875				880	
Leu	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln
			885						890					895	
Ile	His	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe
			900					905					910		
Tyr	Pro	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr
		915					920					925			
Phe	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg
	930					935					940				
Phe	Ala	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn
945					950					955					960
Phe	Glu	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu
			965						970					975	
Arg	Met	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro
			980					985					990		
Lys	His	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr
		995					1000					1005			
Lys	Val	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	
	1010					1015					1020				
Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	
	1025					1030					1035				
Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	
	1040					1045					1050				
Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	
	1055					1060					1065				
Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	
	1070					1075					1080				
Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	
	1085					1090					1095				
Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	
	1100					1105					1110				
Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	Asp	
	1115					1120					1125				
Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	
	1130					1135					1140				
Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	
	1145					1150					1155				
Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	
	1160					1165					1170				
Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	
	1175					1180					1185				
Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	
	1190					1195					1200				
His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	
	1205					1210					1215				
Gly	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	
	1220					1225					1230				

-continued

Met Gly 1235	Arg	His	Lys	Pro	Glu 1240	Asn	Ile	Val	Ile	Glu 1245	Met	Ala	Arg
Glu Asn 1250	Gln	Thr	Thr	Gln	Lys 1255	Gly	Gln	Lys	Asn	Ser 1260	Arg	Glu	Arg
Met Lys 1265	Arg	Ile	Glu	Glu	Gly 1270	Ile	Lys	Glu	Leu	Gly 1275	Ser	Gln	Ile
Leu Lys 1280	Glu	His	Pro	Val	Glu 1285	Asn	Thr	Gln	Leu	Gln 1290	Asn	Glu	Lys
Leu Tyr 1295	Leu	Tyr	Tyr	Leu	Gln 1300	Asn	Gly	Arg	Asp	Met 1305	Tyr	Val	Asp
Gln Glu 1310	Leu	Asp	Ile	Asn	Arg 1315	Leu	Ser	Asp	Tyr	Asp 1320	Val	Asp	Ala
Ile Val 1325	Pro	Gln	Ser	Phe	Leu 1330	Lys	Asp	Asp	Ser	Ile 1335	Asp	Asn	Lys
Val Leu 1340	Thr	Arg	Ser	Asp	Lys 1345	Asn	Arg	Gly	Lys	Ser 1350	Asp	Asn	Val
Pro Ser 1355	Glu	Glu	Val	Val	Lys 1360	Lys	Met	Lys	Asn	Tyr 1365	Trp	Arg	Gln
Leu Leu 1370	Asn	Ala	Lys	Leu	Ile 1375	Thr	Gln	Arg	Lys	Phe 1380	Asp	Asn	Leu
Thr Lys 1385	Ala	Glu	Arg	Gly	Gly 1390	Leu	Ser	Glu	Leu	Asp 1395	Lys	Ala	Gly
Phe Ile 1400	Lys	Arg	Gln	Leu	Val 1405	Glu	Thr	Arg	Gln	Ile 1410	Thr	Lys	His
Val Ala 1415	Gln	Ile	Leu	Asp	Ser 1420	Arg	Met	Asn	Thr	Lys 1425	Tyr	Asp	Glu
Asn Asp 1430	Lys	Leu	Ile	Arg	Glu 1435	Val	Lys	Val	Ile	Thr 1440	Leu	Lys	Ser
Lys Leu 1445	Val	Ser	Asp	Phe	Arg 1450	Lys	Asp	Phe	Gln	Phe 1455	Tyr	Lys	Val
Arg Glu 1460	Ile	Asn	Asn	Tyr	His 1465	His	Ala	His	Asp	Ala 1470	Tyr	Leu	Asn
Ala Val 1475	Val	Gly	Thr	Ala	Leu 1480	Ile	Lys	Lys	Tyr	Pro 1485	Lys	Leu	Glu
Ser Glu 1490	Phe	Val	Tyr	Gly	Asp 1495	Tyr	Lys	Val	Tyr	Asp 1500	Val	Arg	Lys
Met Ile 1505	Ala	Lys	Ser	Glu	Gln 1510	Glu	Ile	Gly	Lys	Ala 1515	Thr	Ala	Lys
Tyr Phe 1520	Phe	Tyr	Ser	Asn	Ile 1525	Met	Asn	Phe	Phe	Lys 1530	Thr	Glu	Ile
Thr Leu 1535	Ala	Asn	Gly	Glu	Ile 1540	Arg	Lys	Arg	Pro	Leu 1545	Ile	Glu	Thr
Asn Gly 1550	Glu	Thr	Gly	Glu	Ile 1555	Val	Trp	Asp	Lys	Gly 1560	Arg	Asp	Phe
Ala Thr 1565	Val	Arg	Lys	Val	Leu 1570	Ser	Met	Pro	Gln	Val 1575	Asn	Ile	Val
Lys Lys 1580	Thr	Glu	Val	Gln	Thr 1585	Gly	Gly	Phe	Ser	Lys 1590	Glu	Ser	Ile
Leu Pro 1595	Lys	Arg	Asn	Ser	Asp 1600	Lys	Leu	Ile	Ala	Arg 1605	Lys	Lys	Asp
Trp Asp	Pro	Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala

-continued

1610	1615	1620
Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys		
1625	1630	1635
Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu		
1640	1645	1650
Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys		
1655	1660	1665
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys		
1670	1675	1680
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala		
1685	1690	1695
Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser		
1700	1705	1710
Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu		
1715	1720	1725
Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu		
1730	1735	1740
Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu		
1745	1750	1755
Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val		
1760	1765	1770
Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln		
1775	1780	1785
Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala		
1790	1795	1800
Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg		
1805	1810	1815
Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln		
1820	1825	1830
Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu		
1835	1840	1845
Gly Gly Asp		
1850		

<210> SEQ ID NO 36

<211> LENGTH: 1846

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 36

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 Asp Lys
 20 25 30

Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala
 35 40 45

Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu
 50 55 60

Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu
 65 70 75 80

Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr

-continued

85								90					95				
Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys	Tyr	Leu	Gln		
			100					105					110				
Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser	Phe	Phe	His		
		115					120					125					
Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg		
	130					135					140						
His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys		
145					150					155					160		
Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp		
			165						170					175			
Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys		
		180						185					190				
Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser		
		195					200					205					
Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu		
	210					215					220						
Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile		
225					230					235					240		
Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala		
			245						250					255			
Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala		
			260					265					270				
Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala		
		275					280					285					
Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu		
	290					295					300						
Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu		
305					310					315					320		
Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg		
			325						330					335			
Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys		
			340					345					350				
Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val		
	355						360					365					
Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser		
	370					375					380						
Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu		
385					390					395					400		
Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu		
			405						410					415			
Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg		
			420					425					430				
Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu		
		435					440					445					
His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp		
	450					455					460						
Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr		
465					470					475					480		
Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg		
			485						490					495			

-continued

Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp	
			500					505					510			
Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp	
		515					520					525				
Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	
		530				535					540					
Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr	
545					550					555					560	
Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	
				565					570					575		
Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	
		580						585					590			
Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	
		595					600					605				
Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	
	610					615					620					
Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	
625					630					635					640	
Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	
			645						650					655		
Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	
		660						665					670			
Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	
		675					680					685				
Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	
	690					695					700					
Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	
705					710					715					720	
Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	
			725						730					735		
Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	
		740						745					750			
Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	
		755					760					765				
Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys	
	770					775					780					
Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr	Gln	
785					790					795					800	
Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu	Glu	Gly	
			805						810					815		
Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	Val	Glu	Asn	
		820						825					830			
Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	Gln	Asn	Gly	
		835					840					845				
Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	Leu	Ser	Asp	
	850					855					860					
Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	Asp	Asp	Ser	
865					870					875					880	
Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	Gly	Lys	Ser	
			885						890					895		

-continued

Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp	900	905	910
Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	Phe	Asp	Asn	915	920	925
Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	Lys	Ala	Gly	930	935	940
Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	Lys	His	Val	945	950	955
Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	Glu	Asn	Asp	965	970	975
Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	Lys	Leu	Val	980	985	990
Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	Glu	Ile	Asn	995	1000	1005
Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	Val	Gly		1010	1015	1020
Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	Val		1025	1030	1035
Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala	Lys		1040	1045	1050
Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe	Tyr		1055	1060	1065
Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala	Asn		1070	1075	1080
Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu	Thr		1085	1090	1095
Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val	Arg		1100	1105	1110
Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr	Glu		1115	1120	1125
Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys	Arg		1130	1135	1140
Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro	Lys		1145	1150	1155
Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val	Leu		1160	1165	1170
Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys	Ser		1175	1180	1185
Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser	Phe		1190	1195	1200
Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys	Glu		1205	1210	1215
Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu	Phe		1220	1225	1230
Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly	Glu		1235	1240	1245
Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val	Asn		1250	1255	1260
Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser	Pro		1265	1270	1275
Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys	His				

-continued

1280	1285	1290
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg 1295 1300 1305		
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr 1310 1315 1320		
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile 1325 1330 1335		
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe 1340 1345 1350		
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr 1355 1360 1365		
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly 1370 1375 1380		
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Gly 1385 1390 1395		
Gly Gly Gly Ser Gly Gly Gly Gly Ser Ser Met Gly Thr Gly Thr 1400 1405 1410		
Lys Cys Ile Gly Gln Ser Lys Met Arg Lys Asn Gly Asp Ile Leu 1415 1420 1425		
Asn Asp Ser His Ala Glu Val Ile Ala Arg Arg Ser Phe Gln Arg 1430 1435 1440		
Tyr Leu Leu His Gln Leu Gln Leu Ala Ala Thr Leu Lys Glu Asp 1445 1450 1455		
Ser Ile Phe Val Pro Gly Thr Gln Lys Gly Val Trp Lys Leu Arg 1460 1465 1470		
Arg Asp Leu Ile Phe Val Phe Phe Ser Ser His Thr Pro Cys Gly 1475 1480 1485		
Asp Ala Ser Ile Ile Pro Met Leu Glu Phe Glu Asp Gln Pro Cys 1490 1495 1500		
Cys Pro Val Phe Arg Asn Trp Ala His Asn Ser Ser Val Glu Ala 1505 1510 1515		
Ser Ser Asn Leu Glu Ala Pro Gly Asn Glu Arg Lys Cys Glu Asp 1520 1525 1530		
Pro Asp Ser Pro Val Thr Lys Lys Met Arg Leu Glu Pro Gly Thr 1535 1540 1545		
Ala Ala Arg Glu Val Thr Asn Gly Ala Ala His His Gln Ser Phe 1550 1555 1560		
Gly Lys Gln Lys Ser Gly Pro Ile Ser Pro Gly Ile His Ser Cys 1565 1570 1575		
Asp Leu Thr Val Glu Gly Leu Ala Thr Val Thr Arg Ile Ala Pro 1580 1585 1590		
Gly Ser Ala Lys Val Ile Asp Val Tyr Arg Thr Gly Ala Lys Cys 1595 1600 1605		
Val Pro Gly Glu Ala Gly Asp Ser Gly Lys Pro Gly Ala Ala Phe 1610 1615 1620		
His Gln Val Gly Leu Leu Arg Val Lys Pro Gly Arg Gly Asp Arg 1625 1630 1635		
Thr Arg Ser Met Ser Cys Ser Asp Lys Met Ala Arg Trp Asn Val 1640 1645 1650		
Leu Gly Cys Gln Gly Ala Leu Leu Met His Leu Leu Glu Glu Pro 1655 1660 1665		

-continued

```

Ile Tyr  Leu Ser Ala Val Val  Ile Gly Lys Cys Pro  Tyr Ser Gln
1670                                1675                                1680

Glu Ala  Met Gln Arg Ala Leu  Ile Gly Arg Cys Gln  Asn Val Ser
1685                                1690                                1695

Ala Leu  Pro Lys Gly Phe Gly  Val Gln Glu Leu Lys  Ile Leu Gln
1700                                1705                                1710

Ser Asp  Leu Leu Phe Glu Gln  Ser Arg Ser Ala Val  Gln Ala Lys
1715                                1720                                1725

Arg Ala  Asp Ser Pro Gly Arg  Leu Val Pro Cys Gly  Ala Ala Ile
1730                                1735                                1740

Ser Trp  Ser Ala Val Pro Glu  Gln Pro Leu Asp Val  Thr Ala Asn
1745                                1750                                1755

Gly Phe  Pro Gln Gly Thr Thr  Lys Lys Thr Ile Gly  Ser Leu Gln
1760                                1765                                1770

Ala Arg  Ser Gln Ile Ser Lys  Val Glu Leu Phe Arg  Ser Phe Gln
1775                                1780                                1785

Lys Leu  Leu Ser Arg Ile Ala  Arg Asp Lys Trp Pro  His Ser Leu
1790                                1795                                1800

Arg Val  Gln Lys Leu Asp Thr  Tyr Gln Glu Tyr Lys  Glu Ala Ala
1805                                1810                                1815

Ser Ser  Tyr Gln Glu Ala Trp  Ser Thr Leu Arg Lys  Gln Val Phe
1820                                1825                                1830

Gly Ser  Trp Ile Arg Asn Pro  Pro Asp Tyr His Gln  Phe
1835                                1840                                1845

```

```

<210> SEQ ID NO 37
<211> LENGTH: 1368
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

```

```

<400> SEQUENCE: 37

```

```

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 165	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro
Asp	Asn	Ser	Asp 180	Val	Asp	Lys	Leu	Phe 185	Ile	Gln	Leu	Val	Gln 190	Thr	Tyr
Asn	Gln	Leu	Phe 195	Glu	Glu	Asn	Pro 200	Ile	Asn	Ala	Ser	Gly 205	Val	Asp	Ala
Lys	Ala	Ile	Leu 210	Ser	Ala	Arg 215	Leu	Ser	Lys	Ser	Arg 220	Arg	Leu	Glu	Asn
Leu 225	Ile	Ala	Gln	Leu	Pro 230	Gly	Glu	Lys	Lys	Asn 235	Gly	Leu	Phe	Gly	Asn 240
Leu	Ile	Ala	Leu	Ser 245	Leu	Gly	Leu	Thr	Pro 250	Asn	Phe	Lys	Ser	Asn 255	Phe
Asp	Leu	Ala	Glu 260	Asp	Ala	Lys	Leu	Gln 265	Leu	Ser	Lys	Asp	Thr 270	Tyr	Asp
Asp	Asp	Leu	Asp 275	Asn	Leu	Leu	Ala 280	Gln	Ile	Gly	Asp	Gln 285	Tyr	Ala	Asp
Leu	Phe 290	Leu	Ala	Ala	Lys	Asn 295	Leu	Ser	Asp	Ala	Ile 300	Leu	Leu	Ser	Asp
Ile 305	Leu	Arg	Val	Asn	Thr 310	Glu	Ile	Thr	Lys	Ala 315	Pro	Leu	Ser	Ala	Ser 320
Met	Ile	Lys	Arg 325	Tyr	Asp	Glu	His	His	Gln 330	Asp	Leu	Thr	Leu 335	Leu	Lys
Ala	Leu	Val	Arg 340	Gln	Gln	Leu	Pro	Glu 345	Lys	Tyr	Lys	Glu	Ile 350	Phe	Phe
Asp	Gln	Ser 355	Lys	Asn	Gly	Tyr	Ala 360	Gly	Tyr	Ile	Asp	Gly 365	Gly	Ala	Ser
Gln 370	Glu	Glu	Phe	Tyr	Lys	Phe 375	Ile	Lys	Pro	Ile	Leu 380	Glu	Lys	Met	Asp
Gly 385	Thr	Glu	Glu	Leu	Leu 390	Val	Lys	Leu	Asn	Arg 395	Glu	Asp	Leu	Leu	Arg 400
Lys	Gln	Arg	Thr 405	Phe	Asp	Asn	Gly	Ser	Ile 410	Pro	His	Gln	Ile 415	His	Leu
Gly	Glu	Leu	His 420	Ala	Ile	Leu	Arg	Arg 425	Gln	Glu	Asp	Phe	Tyr 430	Pro	Phe
Leu	Lys	Asp 435	Asn	Arg	Glu	Lys	Ile 440	Glu	Lys	Ile	Leu 445	Thr	Phe	Arg	Ile
Pro	Tyr 450	Tyr	Val	Gly	Pro	Leu 455	Ala	Arg	Gly	Asn	Ser 460	Arg	Phe	Ala	Trp
Met 465	Thr	Arg	Lys	Ser	Glu 470	Glu	Thr	Ile	Thr	Pro 475	Trp	Asn	Phe	Glu	Glu 480
Val	Val	Asp	Lys 485	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met 495	Thr
Asn	Phe	Asp 500	Lys	Asn	Leu	Pro	Asn	Glu 505	Lys	Val	Leu	Pro	Lys 510	His	Ser
Leu	Leu	Tyr 515	Glu	Tyr	Phe	Thr	Val 520	Tyr	Asn	Glu	Leu	Thr 525	Lys	Val	Lys
Tyr 530	Val	Thr	Glu	Gly	Met	Arg 535	Lys	Pro	Ala	Phe	Leu	Ser 540	Gly	Glu	Gln
Lys 545	Lys	Ala	Ile	Val	Asp 550	Leu	Leu	Phe	Lys	Thr	Asn 555	Arg	Lys	Val	Thr 560

Val	Lys	Gln	Leu	Lys 565	Glu	Asp	Tyr	Phe	Lys 570	Lys	Ile	Glu	Cys	Phe	Asp 575
Ser	Val	Glu	Ile 580	Ser	Gly	Val	Glu	Asp 585	Arg	Phe	Asn	Ala	Ser 590	Leu	Gly
Thr	Tyr	His 595	Asp	Leu	Leu	Lys	Ile 600	Ile	Lys	Asp	Lys	Asp 605	Phe	Leu	Asp
Asn	Glu	Glu	Asn	Glu	Asp	Ile 615	Leu	Glu	Asp	Ile	Val 620	Leu	Thr	Leu	Thr
Leu	Phe	Glu	Asp	Arg	Glu 630	Met	Ile	Glu	Glu	Arg 635	Leu	Lys	Thr	Tyr	Ala 640
His	Leu	Phe	Asp	Asp 645	Lys	Val	Met	Lys	Gln 650	Leu	Lys	Arg	Arg	Arg	Tyr 655
Thr	Gly	Trp	Gly 660	Arg	Leu	Ser	Arg	Lys 665	Leu	Ile	Asn	Gly	Ile 670	Arg	Asp
Lys	Gln	Ser	Gly 675	Lys	Thr	Ile	Leu 680	Asp	Phe	Leu	Lys	Ser 685	Asp	Gly	Phe
Ala	Asn	Arg	Asn	Phe	Met 695	Gln	Leu	Ile	His	Asp	Asp 700	Ser	Leu	Thr	Phe
Lys	Glu	Asp	Ile	Gln	Lys 710	Ala	Gln	Val	Ser	Gly 715	Gln	Gly	Asp	Ser	Leu 720
His	Glu	His	Ile	Ala	Asn 725	Leu	Ala	Gly	Ser 730	Pro	Ala	Ile	Lys	Lys	Gly 735
Ile	Leu	Gln	Thr 740	Val	Lys	Val	Val	Asp 745	Glu	Leu	Val	Lys	Val 750	Met	Gly
Arg	His	Lys 755	Pro	Glu	Asn	Ile	Val 760	Ile	Glu	Met	Ala	Arg 765	Glu	Asn	Gln
Thr	Thr 770	Gln	Lys	Gly	Gln	Lys 775	Asn	Ser	Arg	Glu	Arg 780	Met	Lys	Arg	Ile
Glu	Glu	Gly	Ile	Lys	Glu 790	Leu	Gly	Ser	Gln	Ile 795	Leu	Lys	Glu	His	Pro 800
Val	Glu	Asn	Thr 805	Gln	Leu	Gln	Asn	Glu	Lys 810	Leu	Tyr	Leu	Tyr	Tyr	Leu
Gln	Asn	Gly	Arg 820	Asp	Met	Tyr	Val	Asp 825	Gln	Glu	Leu	Asp 830	Ile	Asn	Arg
Leu	Ser	Asp 835	Tyr	Asp	Val	Asp	Ala 840	Ile	Val	Pro	Gln	Ser 845	Phe	Leu	Lys
Asp	Asp 850	Ser	Ile	Asp	Asn	Lys 855	Val	Leu	Thr	Arg	Ser 860	Asp	Lys	Asn	Arg
Gly	Lys	Ser	Asp	Asn	Val 870	Pro	Ser	Glu	Glu	Val 875	Val	Lys	Lys	Met	Lys 880
Asn	Tyr	Trp	Arg 885	Gln	Leu	Leu	Asn	Ala	Lys 890	Leu	Ile	Thr	Gln	Arg	Lys 895
Phe	Asp	Asn	Leu 900	Thr	Lys	Ala	Glu	Arg 905	Gly	Gly	Leu	Ser	Glu 910	Leu	Asp
Lys	Ala	Gly 915	Phe	Ile	Lys	Arg	Gln 920	Leu	Val	Glu	Thr	Arg 925	Gln	Ile	Thr
Lys	His 930	Val	Ala	Gln	Ile	Leu	Asp 935	Ser	Arg	Met	Asn 940	Thr	Lys	Tyr	Asp
Glu	Asn	Asp	Lys	Leu	Ile 950	Arg	Glu	Val	Lys	Val 955	Ile	Thr	Leu	Lys	Ser 960
Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg

-continued

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			

-continued

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

<210> SEQ ID NO 38
 <211> LENGTH: 82
 <212> TYPE: RNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 38

guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu ccguuaucaa cuugaaaaag 60

uggcaccgag ucgugcuuu uu 82

<210> SEQ ID NO 39
 <211> LENGTH: 180
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 39

gatgacattg catacattcg aaagacccta gccttagata aaactgagca agaggctttg 60

gagtatttca tgaacacaaat gaatgatgca cgtcatggtg gctggacaac aaaaatggat 120

tggatcttcc acacaattaa acagcatgca ttgaactgaa agataactga gaaaatgaaa 180

<210> SEQ ID NO 40
 <211> LENGTH: 59
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 40

Asp Asp Ile Ala Tyr Ile Arg Lys Thr Leu Ala Leu Asp Lys Thr Glu
 1 5 10 15

Gln Glu Ala Leu Glu Tyr Phe Met Lys Gln Met Asn Asp Ala Arg His
 20 25 30

Gly Gly Trp Thr Thr Lys Met Asp Trp Ile Phe His Thr Ile Lys Gln
 35 40 45

His Ala Leu Asn Lys Ile Thr Glu Lys Met Lys
 50 55

<210> SEQ ID NO 41
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 41

aucggaauct auuuugacuc 20

<210> SEQ ID NO 42
 <211> LENGTH: 20
 <212> TYPE: RNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic Polynucleotide

-continued

<400> SEQUENCE: 42

ucggaaucua uuuugacucg 20

<210> SEQ ID NO 43

<211> LENGTH: 20

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 43

cuuagauaaa acugagcaag 20

<210> SEQ ID NO 44

<211> LENGTH: 20

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 44

aucuaauuug acucguucuc 20

<210> SEQ ID NO 45

<211> LENGTH: 20

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 45

uaaaacugag caagaggcuu 20

<210> SEQ ID NO 46

<211> LENGTH: 20

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 46

ugguggcugg acaacaaaaa 20

<210> SEQ ID NO 47

<211> LENGTH: 20

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 47

gcuggacaac aaaauggau 20

<210> SEQ ID NO 48

<211> LENGTH: 20

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 48

guguuaauuu gucguacgua 20

-continued

<210> SEQ ID NO 49
<211> LENGTH: 180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 49

```
aatcacattt ttccacttct tgaaaagtac tgtggcttcc atgaagataa cattccccag      60
ctggaagacg tttctcaatt cctgcagact tgcactggtc tccgcctccg acctgtggct      120
ggcctgcttt cctctcgga tttcttgggt ggcctggcct tccgagtctt ccactgcaca      180
```

<210> SEQ ID NO 50
<211> LENGTH: 60
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 50

```
Asn His Ile Phe Pro Leu Leu Glu Lys Tyr Cys Gly Phe His Glu Asp
1           5           10          15
Asn Ile Pro Gln Leu Glu Asp Val Ser Gln Phe Leu Gln Thr Cys Thr
          20          25          30
Gly Ser Arg Leu Arg Pro Val Ala Gly Leu Leu Ser Ser Arg Asp Phe
          35          40          45
Leu Gly Gly Leu Ala Phe Arg Val Phe His Cys Thr
          50          55          60
```

<210> SEQ ID NO 51
<211> LENGTH: 180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 51

```
atgcctgcct ggggagccct gttcctgctc tggggccacag cagaggccac caaggactgc      60
cccagcccac gtacctgccg cgccctggaa accatggggc tgtgggtgga ctgcaggggc      120
cacggactca cggccctgcc tgccctgccg gcccgacccc gccaccttct gctggccaac      180
```

<210> SEQ ID NO 52
<211> LENGTH: 60
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 52

```
Met Pro Ala Trp Gly Ala Leu Phe Leu Leu Trp Ala Thr Ala Glu Ala
1           5           10          15
Thr Lys Asp Cys Pro Ser Pro Arg Thr Cys Arg Ala Leu Glu Thr Met
          20          25          30
Gly Leu Trp Val Asp Cys Arg Gly His Gly Leu Thr Ala Leu Pro Ala
          35          40          45
Leu Pro Ala Arg Thr Arg His Leu Leu Leu Ala Asn
          50          55          60
```

-continued

<210> SEQ ID NO 53
<211> LENGTH: 120
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 53

```
ggttatggtc ctgtctgccc tcttggtggc atacaagaag tcaatatcaa ccagagccct    60
cttcagcccc tcaatgtgga gattgaccct gagatccaaa aggtgaagtc tcgagaaagg    120
```

<210> SEQ ID NO 54
<211> LENGTH: 40
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 54

```
Gly Tyr Gly Pro Val Cys Pro Pro Gly Gly Ile Gln Glu Val Thr Ile
1           5           10          15
Asn Gln Ser Pro Leu Gln Pro Leu Asn Val Glu Ile Asp Pro Glu Ile
          20          25          30
Gln Lys Val Lys Ser Arg Glu Arg
          35          40
```

<210> SEQ ID NO 55
<211> LENGTH: 180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 55

```
gtctccctgg ctgaggatcc ccaggagatc gctgcccaaga agacagatac atcccaccat    60
gatcaggatc acccaacctt caacaagatc acccccaacc cggctgagtt cgccttcagc    120
ctataccgcc agctggcaca ccagtccaac agcaccaata tcttcttctc cccagtgagc    180
```

<210> SEQ ID NO 56
<211> LENGTH: 60
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 56

```
Val Ser Leu Ala Glu Asp Pro Gln Gly Asp Ala Ala Gln Lys Thr Asp
1           5           10          15
Thr Ser His His Asp Gln Asp His Pro Thr Phe Asn Lys Ile Thr Pro
          20          25          30
Asn Pro Ala Glu Phe Ala Phe Ser Leu Tyr Arg Gln Leu Ala His Gln
          35          40          45
Ser Asn Ser Thr Asn Ile Phe Phe Ser Pro Val Ser
          50          55          60
```

<210> SEQ ID NO 57
<211> LENGTH: 180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 57

```
ggccactgcc tcattatcaa caatgtgaac ttctgccgtg agtcgggct ccgcaccgc      60
actggctcca acatcgactg tgagaagttg cggcgctgct tctctcgcc gcatttcattg    120
gtggaggtga agggcgacct gactgccaaag aaaatggtgc tggctttgct ggagctggcg    180
```

<210> SEQ ID NO 58

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 58

```
Gly His Cys Leu Ile Ile Asn Asn Val Asn Phe Cys Arg Glu Ser Gly
1           5           10          15
Leu Arg Thr Arg Thr Gly Ser Asn Ile Asp Cys Glu Lys Leu Arg Arg
          20          25          30
Arg Phe Ser Ser Pro His Phe Met Val Glu Val Lys Gly Asp Leu Thr
          35          40          45
Ala Lys Lys Met Val Leu Ala Leu Leu Glu Leu Ala
          50          55          60
```

<210> SEQ ID NO 59

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 59

```
actagagcta gatactttct agttgggagc aataatgcag aaacgaaata tcgtgtcttg      60
aagactgata gaacagaacc aaaagatttg gtcataattg atgacaggca tgtctatact    120
caacaagaag taagggaact tcttgccgc ttgatcttg gaaatagaac aaagatggga    180
```

<210> SEQ ID NO 60

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 60

```
Thr Arg Ala Arg Tyr Phe Leu Val Gly Ser Asn Asn Ala Glu Thr Lys
1           5           10          15
Tyr Arg Val Leu Lys Thr Asp Arg Thr Glu Pro Lys Asp Leu Val Ile
          20          25          30
Ile Asp Asp Arg His Val Tyr Thr Gln Gln Glu Val Arg Glu Leu Leu
          35          40          45
Gly Arg Leu Asp Leu Gly Asn Arg Thr Lys Met Gly
          50          55          60
```

<210> SEQ ID NO 61

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 61

```
acagatgccc cggtagagccc caccactctg tatgtggagg acatctcgga accgccgttg      60
cacgatttct accgcagcag gctactggac ctggctcttc tgctggatgg ctctccagg      120
ctgtccgagg ctgagtttga agtgctgaag gcctttgtgg tggacatgat ggagcggctg      180
```

<210> SEQ ID NO 62

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 62

```
Thr Asp Ala Pro Val Ser Pro Thr Thr Leu Tyr Val Glu Asp Ile Ser
1           5           10          15
Glu Pro Pro Leu His Asp Phe Tyr Arg Ser Arg Leu Leu Asp Leu Val
          20          25          30
Phe Leu Leu Asp Gly Ser Ser Arg Leu Ser Glu Ala Glu Phe Glu Val
          35          40          45
Leu Lys Ala Phe Val Val Asp Met Met Glu Arg Leu
          50          55          60
```

<210> SEQ ID NO 63

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 63

```
atctgtgctg ctgtcctcag caaattcatg tctgtgttct gcgggggtata tgagcagcca      60
tactactact ctgatatcct gacggtgggc tgtgctgtgg gactcggccg ttgttttggg      120
acaccacttg gaggagtgcct atttagcatc gaggtcacct ccacctactt tgctgttcgg      180
```

<210> SEQ ID NO 64

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 64

```
Ile Cys Ala Ala Val Leu Ser Lys Phe Met Ser Val Phe Cys Gly Val
1           5           10          15
Tyr Glu Gln Pro Tyr Tyr Tyr Ser Asp Ile Leu Thr Val Gly Cys Ala
          20          25          30
Val Gly Val Gly Arg Cys Phe Gly Thr Pro Leu Gly Gly Val Leu Phe
          35          40          45
Ser Ile Glu Val Thr Ser Thr Tyr Phe Ala Val Arg
          50          55          60
```

<210> SEQ ID NO 65

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 65

```
tactttgaaa agtcaaagga gcagctgaca cccctgatca agaaggctgg aacggaactg    60
gttaacttct tgagctatatt cgtggaactt ggaacacagc ctgccaccca gcgaagtgtc    120
cagcaccatt gtcttccaac cccagctggc ctctagaaca cccactggcc agtcctagag    180
```

<210> SEQ ID NO 66

<211> LENGTH: 59

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 66

```
Tyr Phe Glu Lys Ser Lys Glu Gln Leu Thr Pro Leu Ile Lys Lys Ala
1             5             10             15
Gly Thr Glu Leu Val Asn Phe Leu Ser Tyr Phe Val Glu Leu Gly Thr
20            25            30
Gln Pro Ala Thr Gln Arg Ser Val Gln His His Cys Leu Pro Thr Pro
35            40            45
Ala Gly Leu Asn Thr His Trp Pro Val Leu Glu
50            55
```

<210> SEQ ID NO 67

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 67

```
ccgcacaagc gcctcacgct cagcggcatc tgcgccttca ttagtgaccg cttcccctac    60
taccgcgcga agttccccgc ccggcagaac agcatccgcc acaacctctc gctgaacgac    120
tgcttcgtca agatcccccg cgagccgggc cgcccaggca agggcaacta ctggagcctg    180
```

<210> SEQ ID NO 68

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 68

```
Pro His Lys Arg Leu Thr Leu Ser Gly Ile Cys Ala Phe Ile Ser Asp
1             5             10             15
Arg Phe Pro Tyr Tyr Arg Arg Lys Phe Pro Ala Arg Gln Asn Ser Ile
20            25            30
Arg His Asn Leu Ser Leu Asn Asp Cys Phe Val Lys Ile Pro Arg Glu
35            40            45
Pro Gly Arg Pro Gly Lys Gly Asn Tyr Trp Ser Leu
50            55            60
```

<210> SEQ ID NO 69

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 69

```
gctgaggacc tgtggctgag cccgctgacc atggaagatc ttgtctgcta cagcttcag      60
gtggccagag ggatggagtt cctggcttcc cgaaagtgca tccgcagaga cctggctgct      120
cggaacattc tgctgtcgga aagcgacgtg gtgaagatct gtgacttttg ccttgcccgg      180
```

<210> SEQ ID NO 70

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 70

```
Ala Glu Asp Leu Trp Leu Ser Pro Leu Thr Met Glu Asp Leu Val Cys
 1             5             10            15
Tyr Ser Phe Gln Val Ala Arg Gly Met Glu Phe Leu Ala Ser Arg Lys
      20            25            30
Cys Ile Arg Arg Asp Leu Ala Ala Arg Asn Ile Leu Leu Ser Glu Ser
      35            40            45
Asp Val Val Lys Ile Cys Asp Phe Gly Leu Ala Arg
      50            55            60
```

<210> SEQ ID NO 71

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 71

```
gataccgaga ctgtggggcca gagagccctg cactcaattc tgaatgctgc catcatgac      60
agtgtcgttg ttgtcatgac tatcctcctg gtggttctgt ataaatacag gtgtataag      120
gtcatccatg cctggcttat tatatcatct ctattgttgc tgttcttttt ttcattcatt      180
```

<210> SEQ ID NO 72

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 72

```
Asp Thr Glu Thr Val Gly Gln Arg Ala Leu His Ser Ile Leu Asn Ala
 1             5             10            15
Ala Ile Met Ile Ser Val Val Val Met Thr Ile Leu Leu Val Val
      20            25            30
Leu Tyr Lys Tyr Arg Cys Tyr Lys Val Ile His Ala Trp Leu Ile Ile
      35            40            45
Ser Ser Leu Leu Leu Leu Phe Phe Phe Ser Phe Ile
      50            55            60
```

<210> SEQ ID NO 73

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 73

```
aagccgagta agccaaaaac caacatgaag cacatggctg gtgctgcagc agctggggca    60
gtggtggggg gccttgccgg ctacgtgctg ggaagtgcc tgcagggcc catcatacat    120
ttcggcagtg actatgagga ccgttactat cgtgaaaaca tgcaccgtta cccaaccaa    180
```

<210> SEQ ID NO 74

<211> LENGTH: 60

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 74

```
Lys Pro Ser Lys Pro Lys Thr Asn Met Lys His Met Ala Gly Ala Ala
1           5           10          15
Ala Ala Gly Ala Val Val Gly Gly Leu Gly Gly Tyr Val Leu Gly Ser
20          25          30
Ala Met Ser Arg Pro Ile Ile His Phe Gly Ser Asp Tyr Glu Asp Arg
35          40          45
Tyr Tyr Arg Glu Asn Met His Arg Tyr Pro Asn Gln
50          55          60
```

<210> SEQ ID NO 75

<211> LENGTH: 120

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 75

```
cttcccagcc gagacgtgac agtccttctg gaaaactatg gcaaattcga aaaggggtgt    60
ttgatttttg ttgtacgttt cctctttggc ctggtaaacc aggagaggac ctctacttg    120
```

<210> SEQ ID NO 76

<211> LENGTH: 40

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 76

```
Leu Pro Ser Arg Asp Val Thr Val Leu Leu Glu Asn Tyr Gly Lys Phe
1           5           10          15
Glu Lys Gly Cys Leu Ile Phe Val Val Arg Phe Leu Phe Gly Leu Val
20          25          30
Asn Gln Glu Arg Thr Ser Tyr Leu
35          40
```

<210> SEQ ID NO 77

<211> LENGTH: 180

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 77

-continued

gtgaagcact tctccccaga ggaactcaaa gttaaggtgt tgggagatgt gattgaggtg 60
catggaaaac atgaagagcg ccaggatgaa catggtttca tctccaggga gttccacggg 120
aaataccgga tcccagctga tgtagacct ctcaccatta cttcatccct gtcattctgat 180

<210> SEQ ID NO 78
<211> LENGTH: 60
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 78

Val Lys His Phe Ser Pro Glu Glu Leu Lys Val Lys Val Leu Gly Asp
1 5 10 15
Val Ile Glu Val His Gly Lys His Glu Glu Arg Gln Asp Glu His Gly
20 25 30
Phe Ile Ser Arg Glu Phe His Gly Lys Tyr Arg Ile Pro Ala Asp Val
35 40 45
Asp Pro Leu Thr Ile Thr Ser Ser Leu Ser Ser Asp
50 55 60

<210> SEQ ID NO 79
<211> LENGTH: 180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 79

gagctgcact gtgacaagct gcacgtggat cctgagaact tcaggctcct gggcaacgtg 60
ctggtctgtg tgccggccca tcactttggc aaagaattca cccaccagt gcaggctgcc 120
tatcagaaa tggtggctgg tgtggctaata gccctggccc acaagtatca ctaagctcgc 180

<210> SEQ ID NO 80
<211> LENGTH: 59
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 80

Glu Leu His Cys Asp Lys Leu His Val Asp Pro Glu Asn Phe Arg Leu
1 5 10 15
Leu Gly Asn Val Leu Val Cys Val Pro Ala His His Phe Gly Lys Glu
20 25 30
Phe Thr Pro Pro Val Gln Ala Ala Tyr Gln Lys Val Val Ala Gly Val
35 40 45
Ala Asn Ala Leu Ala His Lys Tyr His Ala Arg
50 55

<210> SEQ ID NO 81
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 81

-continued

aucggaauuct auuuugacuc guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu 60

ccguuaucua cuugaaaaag uggcaccgag ucgguuguuu uu 102

<210> SEQ ID NO 82

<211> LENGTH: 102

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 82

ucggaaucau uuugacucg guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu 60

ccguuaucua cuugaaaaag uggcaccgag ucgguuguuu uu 102

<210> SEQ ID NO 83

<211> LENGTH: 102

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 83

cuuagauaaa acugagcaag guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu 60

ccguuaucua cuugaaaaag uggcaccgag ucgguuguuu uu 102

<210> SEQ ID NO 84

<211> LENGTH: 102

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 84

aucuaauuuug acucguucuc guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu 60

ccguuaucua cuugaaaaag uggcaccgag ucgguuguuu uu 102

<210> SEQ ID NO 85

<211> LENGTH: 102

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 85

uaaaacugag caagaggcuu guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu 60

ccguuaucua cuugaaaaag uggcaccgag ucgguuguuu uu 102

<210> SEQ ID NO 86

<211> LENGTH: 102

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Polynucleotide

<400> SEQUENCE: 86

ugguggcugg acaacaaaaa guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu 60

ccguuaucua cuugaaaaag uggcaccgag ucgguuguuu uu 102

<210> SEQ ID NO 87

-continued

```

<211> LENGTH: 102
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

```

```

<400> SEQUENCE: 87

```

```

gcuggacaac aaaaauggau guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu      60
ccguuaucaa cuugaaaaag uggcaccgag ucgugucuuu uu                          102

```

```

<210> SEQ ID NO 88
<211> LENGTH: 102
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

```

```

<400> SEQUENCE: 88

```

```

guguuaauuu gucguacgua guuuuagagc uagaaauagc aaguuaaaau aaaggcuagu      60
ccguuaucaa cuugaaaaag uggcaccgag ucgugucuuu uu                          102

```

```

<210> SEQ ID NO 89
<211> LENGTH: 180
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polynucleotide

```

```

<400> SEQUENCE: 89

```

```

gcctccgcca acgtggactt cgctttcagc ctgtacaagc agttagtcct gaaggccctt      60
gataagaatg tcatcttctc cccaccgagc atctccaccg ccttggcctt cctgtctctg    120
ggggcccata ataccacctt gacagagatt ctcaaaggcc tcaagttcta cctcacggag    180

```

```

<210> SEQ ID NO 90
<211> LENGTH: 60
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

```

```

<400> SEQUENCE: 90

```

```

Ala Ser Ala Asn Val Asp Phe Ala Phe Ser Leu Tyr Lys Gln Leu Val
1          5          10         15
Leu Lys Ala Pro Ala Ser Ala Asn Val Asp Phe Ala Phe Ser Leu Tyr
20        25        30
Lys Gln Leu Val Leu Lys Ala Pro Gly Ala His Asn Thr Thr Leu Thr
35        40        45
Glu Ile Leu Lys Gly Leu Lys Phe Tyr Leu Thr Glu
50        55        60

```

```

<210> SEQ ID NO 91
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

```

```

<400> SEQUENCE: 91

```

```

Gly Gly Gly Gly Ser
1          5

```

-continued

```

<210> SEQ ID NO 92
<211> LENGTH: 1636
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Polypeptide

<400> SEQUENCE: 92

Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Met Thr Ser Glu Lys
1      5      10      15
Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg Arg Ile Glu Pro Trp
20     25     30
Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu Ala Cys
35     40     45
Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp Arg Ser
50     55     60
Ser Gly Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile Lys Lys
65     70     75     80
Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met Ser Cys Ser Ile Thr
85     90     95
Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg
100    105    110
Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu Val Ile Tyr Val Ala
115    120    125
Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg Gln Gly Leu Arg Asp
130    135    140
Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu Tyr
145    150    155    160
Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp Glu Ala
165    170    175
His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala Leu Glu
180    185    190
Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile Ser Arg
195    200    205
Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu His Leu Gln Asn Cys
210    215    220
His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly Leu Ile
225    230    235    240
His Pro Ser Val Ala Trp Arg Ser Pro Lys Lys Lys Arg Lys Val Glu
245    250    255
Ala Ser Ser Pro Lys Lys Lys Arg Lys Val Glu Ala Ser Asp Lys Lys
260    265    270
Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
275    280    285
Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val Leu Gly
290    295    300
Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu
305    310    315    320
Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
325    330    335
Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu
340    345    350

```

-continued

Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser	Phe	Phe	His	Arg
	355						360					365			
Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	His	Glu	Arg	His
	370					375					380				
Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	His	Glu	Lys	Tyr
385				390						395					400
Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser	Thr	Asp	Lys
			405						410					415	
Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	Met	Ile	Lys	Phe
			420					425					430		
Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp	Asn	Ser	Asp
	435						440					445			
Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr	Asn	Gln	Leu	Phe
	450					455					460				
Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	Lys	Ala	Ile	Leu
465					470					475					480
Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	Leu	Ile	Ala	Gln
			485						490					495	
Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe	Gly	Asn	Leu	Ile	Ala	Leu
			500					505					510		
Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	Asp	Leu	Ala	Glu
	515					520					525				
Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	Asp	Asp	Leu	Asp
	530					535					540				
Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	Leu	Phe	Leu	Ala
545					550					555					560
Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	Ile	Leu	Arg	Val
			565						570					575	
Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	Met	Ile	Lys	Arg
		580						585					590		
Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	Ala	Leu	Val	Arg
	595					600						605			
Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	Asp	Gln	Ser	Lys
	610					615					620				
Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	Gln	Glu	Glu	Phe
625					630					635					640
Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	Gly	Thr	Glu	Glu
			645						650					655	
Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	Lys	Gln	Arg	Thr
		660						665					670		
Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu	Gly	Glu	Leu	His
	675					680						685			
Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	Leu	Lys	Asp	Asn
	690					695					700				
Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	Pro	Tyr	Tyr	Val
705					710					715					720
Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	Met	Thr	Arg	Lys
			725						730					735	
Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	Val	Val	Asp	Lys
		740						745					750		

-continued

Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	Asn	Phe	Asp	Lys	755	760	765
Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser	Leu	Leu	Tyr	Glu	770	775	780
Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys	Tyr	Val	Thr	Glu	785	790	800
Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln	Lys	Lys	Ala	Ile	805	810	815
Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr	Val	Lys	Gln	Leu	820	825	830
Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp	Ser	Val	Glu	Ile	835	840	845
Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly	Thr	Tyr	His	Asp	850	855	860
Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Asn	Glu	Glu	Asn	865	870	875
Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	Leu	Phe	Glu	Asp	885	890	895
Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	His	Leu	Phe	Asp	900	905	910
Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	Thr	Gly	Trp	Gly	915	920	925
Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	Lys	Gln	Ser	Gly	930	935	940
Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	Ala	Asn	Arg	Asn	945	950	955
Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	Lys	Glu	Asp	Ile	965	970	975
Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	His	Glu	His	Ile	980	985	990
Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	Ile	Leu	Gln	Thr	995	1000	1005
Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	Arg	His	Lys		1010	1015	1020
Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	Thr	Thr		1025	1030	1035
Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	Glu		1040	1045	1050
Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro		1055	1060	1065
Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr		1070	1075	1080
Leu	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile		1085	1090	1095
Asn	Arg	Leu	Ser	Asp	Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser		1100	1105	1110
Phe	Leu	Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser		1115	1120	1125
Asp	Lys	Asn	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val		1130	1135	1140
Val	Lys	Lys	Met	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys				

-continued

1145	1150	1155
Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg		
1160	1165	1170
Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln		
1175	1180	1185
Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu		
1190	1195	1200
Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile		
1205	1210	1215
Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp		
1220	1225	1230
Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn		
1235	1240	1245
Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val Gly Thr		
1250	1255	1260
Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val Tyr		
1265	1270	1275
Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys Ser		
1280	1285	1290
Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser		
1295	1300	1305
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly		
1310	1315	1320
Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly		
1325	1330	1335
Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys		
1340	1345	1350
Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val		
1355	1360	1365
Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn		
1370	1375	1380
Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys		
1385	1390	1395
Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val		
1400	1405	1410
Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val		
1415	1420	1425
Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu		
1430	1435	1440
Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val		
1445	1450	1455
Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu		
1460	1465	1470
Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu Leu		
1475	1480	1485
Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn Phe		
1490	1495	1500
Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro Glu		
1505	1510	1515
Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr		
1520	1525	1530

-continued

Leu Asp	Glu Ile Ile Glu Gln	Ile Ser Glu Phe Ser	Lys Arg Val
1535	1540	1545	
Ile Leu	Ala Asp Ala Asn Leu	Asp Lys Val Leu Ser	Ala Tyr Asn
1550	1555	1560	
Lys His	Arg Asp Lys Pro Ile	Arg Glu Gln Ala Glu	Asn Ile Ile
1565	1570	1575	
His Leu	Phe Thr Leu Thr Asn	Leu Gly Ala Pro Ala	Ala Phe Lys
1580	1585	1590	
Tyr Phe	Asp Thr Thr Ile Asp	Arg Lys Arg Tyr Thr	Ser Thr Lys
1595	1600	1605	
Glu Val	Leu Asp Ala Thr Leu	Ile His Gln Ser Ile	Thr Gly Leu
1610	1615	1620	
Tyr Glu	Thr Arg Ile Asp Leu	Ser Gln Leu Gly Gly	Asp
1625	1630	1635	

1. A method of editing a nucleic acid molecule encoding a Caspase-9 protein, the method comprising contacting the nucleic acid molecule with

(a) a fusion protein comprising a nuclease-inactive Cas9 domain and a deaminase domain; and

(b) an sgRNA targeting the fusion protein of (a) to the Caspase-9-encoding nucleic acid molecule;

wherein the nucleic acid molecule comprises a T>C and/or an A>G point mutation in the Caspase-9-encoding nucleic acid molecule as compared to a wild-type Caspase-9-encoding nucleic acid molecule,

and wherein the Caspase-9-encoding nucleic acid molecule is contacted with the fusion protein and the sgRNA in an amount effective and under conditions suitable for the deamination the mutant C or G nucleotide base.

2. The method of claim 1, wherein the deaminase is a cytidine deaminase.

3. The method of claim 1, wherein the deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

4. The method of claim 3, wherein the deaminase is an APOBEC1 family deaminase.

5. The method of claim 3, wherein the deaminase is an activation-induced cytidine deaminase (AID).

6. The method of claim 1, wherein the deaminase is an ACF1/ASE deaminase.

7. The method of claim 1, wherein the deaminase is an adenosine deaminase.

8. The method of claim 7, wherein the deaminase is an ADAT family deaminase.

9. The method of claim 1, wherein the deaminase domain is fused to the N-terminus of the Cas9 domain.

10. The method of claim 1, wherein the deaminase domain is fused to the C-terminus of the Cas9 domain.

11. The method of claim 1, wherein the Cas9 domain and the deaminase domain are fused via a linker.

12. The method of claim 1, wherein the linker comprises a (GGGGS)_n (SEQ ID NO: 91), a (G)_n, an (EAAAK)_n (SEQ

ID NO: 5), or an (XP)_n motif, or a combination of any of these, wherein n is independently an integer between 1 and 30.

13. The method of claim 1, wherein the T>C and/or an A>G point mutation in the Caspase-9-encoding nucleic acid molecule is associated with a disease or disorder, and wherein the deamination of the mutant C or G residue results in a sequence that is not associated with a disease or disorder.

14. The method of claim 1, wherein the deamination results in a correction of the T>C and/or an A>G point mutation to restore the wild-type sequence.

15. The method of claim 1, wherein the sequence associated with the disease or disorder encodes a protein, and wherein the deamination introduces a stop codon into the sequence associated with the disease or disorder, resulting in a truncation of the encoded protein.

16. The method of claim 1, wherein the T>C and/or an A>G point mutation causes an amino acid sequence in the Caspase-9 protein as compared to the wild type Caspase-9 protein.

17. The method of claim 1, wherein the Caspase-9 protein comprises an L197P substitution caused by an T>C point mutation in codon 197 of the Caspase-9 gene.

18. The method of claim 17, wherein the Caspase-9 point mutation is associated with neuroblastoma.

19. The method of claim 17, wherein the contacting results in deamination of the mutant cytidine residue in codon 197 of the Caspase-9 gene, thus correcting the T>C point mutation.

20. The method of claim 1, wherein the contacting is in vivo in a subject having or diagnosed with the disease or disorder.

21. The method of claim 1, wherein the method further comprises detecting the deamination the mutant C or G nucleotide base.

22. The method of claim 21, wherein the detecting is via PCR.

* * * * *