PROGRAMMABLE DNA BASE EDITING BY NME2CAS9-DEAMINASE FUSION PROTEINS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application No. 62/745,666, filed October 15, 2018, herein incorporated by reference in it's entirety.

Field Of The Invention

The present invention is related to the field of gene editing. In particular, the gene editing is directed toward single nucleotide base editing. For example, such single nucleotide base editing results in a conversion of a C•G base pair to a T•A base pair. The high accuracy and precision of the presently disclosed single nucleotide base gene editor is accomplished by an NmeCas9 nuclease that is fused to a nucleotide deaminase protein. The compact nature of the NmeCas9 coupled with a larger number of compatible protospacer adjacent motifs provide the Cas9 fusion constructs contemplated herein to have a gene editing window that can edit sites that are not targetable by other conventional SpyCas9 base editor platforms.

Background

Many human diseases arise due to the mutation of a single base. The ability to correct such genetic aberrations is paramount in treating these genetic disorders. Clustered regularly interspaced short palindromic repeats (CRISPR) along with CRISPR associated (Cas) proteins comprise an RNA-guided adaptive immune system in archaea and bacteria. These systems provide immunity by targeting and inactivating nucleic acids that originate from foreign genetic elements.

SpyCas9 base editing platforms cannot be used to target all single-base mutations due to their limited editing windows. The editing window is constrained in part by the requirement for an NGG PAM and by the requirement that the edited base(s) be a very precise distance from the PAM. SpyCas9 is also intrinsically associated with high off-targeting effects in genome editing.

What is needed in the art is a highly accurate Cas9 single base editing platform having a programmable target specificity due to recognition of a diverse population of PAM sites.

Summary Of The Invention

The present invention is related to the field of gene editing. In particular, the gene editing is directed toward single nucleotide base editing. For example, such single nucleotide base editing results in a conversion of a C•G base pair to a T•A base pair. The high accuracy and precision of the presently disclosed single nucleotide base gene editor is accomplished by an NmeCas9 nuclease that is fused to a nucleotide deaminase protein. The compact nature of the NmeCas9 coupled with a larger number of compatible protospacer adjacent motifs provide the Cas9 fusion constructs contemplated herein to have a gene editing window that is superior to other conventional SpyCas9 base editor platforms.

In one embodiment, the present invention contemplates a mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for an N_4CC nucleotide sequence. In one embodiment, said protein is Nme2Cas9. In one embodiment, said protein further comprises a nuclear localization signal protein. In one embodiment, said nucleotide deaminase is a cytidine deaminase. In one embodiment, said nucleotide deaminase is an adenosine deaminase. In one embodiment, the protein further comprises a uracil glycosylase inhibitor. In one embodiment, the said nuclear localization signal protein includes, but is not limited to, nucleoplasmin (NLS) and/or SV40 NLS and/or C-myc NLS. In one embodiment, said binding region is a protospacer accessory motif interacting domain. In one embodiment, said protospacer accessory motif interacting domain comprises said mutation. In one embodiment, said mutation is a D16A mutation. In one embodiment, said mutated NmeCas9 protein further comprises CBE4. In one embodiment, said mutated NmeCas9 protein further comprises a linker. In one embodiment, said linker is a 73aa linker. In one embodiment, said linker is a 3xHA-tag.

In one embodiment, the present invention contemplates a construct, wherein said construct is an optimized nNme2Cas9-ABEmax.

In one embodiment, the present invention contemplates a construct, wherein said construct is a nNme2Cas9-CBE4.

In one embodiment, the present invention contemplates a construct, wherein said construct is a YE1-BE3-nNme2Cas9 (D16A)-UGI.

In one embodiment, the present invention contemplates an adeno-associated virus comprising a mutated NmeCas9 protein, said mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for an N_4CC nucleotide sequence. In one

embodiment, said virus is an adeno-associated virus 8. In one embodiment, said virus is an adeno-associated virus 6. In one embodiment, said protein is Nme2Cas9. In one embodiment, said protein further comprises a nuclear localization signal protein. In one embodiment, said nucleotide deaminase is a cytidine deaminase. In one embodiment, said nucleotide deaminase is an adenosine deaminase. In one embodiment, the protein further comprises a uracil glycosylase inhibitor. In one embodiment, the nuclear localization signal protein includes, but is not limited to, nucleoplasmin (NLS) and/or SV40 NLS and/or C-myc NLS. In one embodiment, said binding region is a protospacer accessory motif interacting domain. In one embodiment, said protospacer accessory motif interacting domain comprises said mutation. In one embodiment, said mutation is a D16A mutation. In one embodiment, said mutated NmeCas9 protein further comprises CBE4. In one embodiment, said mutated NmeCas9 protein further comprises a linker. In one embodiment, said linker is a 73aa linker. In one embodiment, said linker is a 3xHA-tag.

In one embodiment, the present invention contemplates a construct, wherein said construct is an optimized nNme2Cas9-ABEmax.

In one embodiment, the present invention contemplates a construct, wherein said construct is a nNme2Cas9-CBE4.

In one embodiment, the present invention contemplates a construct, wherein said construct is a YE1-BE3-nNme2Cas9 (D16A)-UGI.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a nucleotide sequence comprising a gene with a mutated single base, wherein said gene is flanked by an N_4CC nucleotide sequence; ii) a mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for said N_4CC nucleotide sequence; b) contacting said nucleotide sequence with said mutated NmeCas9 protein under conditions such that said binding region attaches to said N_4CC nucleotide sequence; and c) replacing said mutated single base with a wild type base with said mutated NmeCas9 protein. In one embodiment, said protein is Nme2Cas9. In one embodiment, said protein further comprises a nuclear localization signal protein. In one embodiment, said nucleotide deaminase is a cytidine deaminase. In one embodiment, said nucleotide deaminase is an adenosine deaminase. In one embodiment, the protein further comprises a uracil glycosylase inhibitor. In one embodiment, the nuclear localization signal protein includes, but is not limited to, nucleoplasmin (NLS) and/or SV40 NLS and/or C-myc NLS. In one embodiment, said binding region is a protospacer

accessory motif interacting domain. In one embodiment, said protospacer accessory motif interacting domain comprises said mutation. In one embodiment, said mutation is a D16A mutation. In one embodiment, said mutated NmeCas9 protein further comprises CBE4. In one embodiment, said mutated NmeCas9 protein further comprises a linker. In one embodiment, said linker is a 73aa linker. In one embodiment, said linker is a 3xHA-tag. In one embodiment, said gene encodes a tyrosinase. In one embodiment, said gene is *Fah.* In one embodiment, said gene is *c-fos*.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a patient comprising a nucleotide sequence comprising a gene with a mutated single base, wherein said gene is flanked by an N_4CC nucleotide sequence, wherein said mutated gene causes a genetically-based medical condition; ii) an adeno-associated virus comprising a mutated NmeCas9 protein, said mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for said N_4CC nucleotide sequence; b) treating said patient with said adeno-associated virus under conditions such that said mutated NmeCas9 protein replaces said mutated single base with a wild type single base, such that said genetically-based medical condition does not develop. In one embodiment, said gene encodes a tyrosinase protein. In one embodiment, said genetically-based medical condition is tyrosinemia. In one embodiment, said virus is an adeno-associated virus 8. In one embodiment, said virus is an adeno-associated virus 6. In one embodiment, said protein is Nme2Cas9. In one embodiment, said protein further comprises a nuclear localization signal protein. In one embodiment, said nucleotide deaminase is a cytidine deaminase. In one embodiment, said nucleotide deaminase is an adenosine deaminase. In one embodiment, the protein further comprises a uracil glycosylase inhibitor. In one embodiment, the nuclear localization signal protein includes, but is not limited to, nucleoplasmin (NLS) and/or SV40 NLS and/or C-myc NLS. In one embodiment, said binding region is a protospacer accessory motif interacting domain. In one embodiment, said protospacer accessory motif interacting domain comprises said mutation. In one embodiment, said mutation is a D16A mutation. In one embodiment, said mutated NmeCas9 protein further comprises CBE4. In one embodiment, said mutated NmeCas9 protein further comprises a linker. In one embodiment, said linker is a 73aa linker. In one embodiment, said linker is a 3xHA-tag. In one embodiment, said gene encodes a tyrosinase. In one embodiment, said gene is *Fah.* In one embodiment, said gene is *c-fos*.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a patient comprising a nucleotide sequence comprising a gene with a mutated single base, wherein said gene is flanked by an N_4CC nucleotide sequence, wherein said mutated gene causes a genetically-based medical condition; ii) an optimized nNme2Cas9-ABEmax, comprising a mutated NmeCas9 protein, said mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for said N_4CC nucleotide sequence; b) treating said patient with said optimized nNme2Cas9-ABEmax under conditions such that said mutated NmeCas9 protein replaces said mutated single base with a wild type single base, such that said genetically-based medical condition does not develop.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a patient comprising a nucleotide sequence comprising a gene with a mutated single base, wherein said gene is flanked by an N_4CC nucleotide sequence, wherein said mutated gene causes a genetically-based medical condition; ii) a nNme2Cas9-CBE4, comprising a mutated NmeCas9 protein, said mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for said N_4CC nucleotide sequence; b) treating said patient with said nNme2Cas9-CBE4 under conditions such that said mutated NmeCas9 protein replaces said mutated single base with a wild type single base, such that said genetically-based medical condition does not develop.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a patient comprising a nucleotide sequence comprising a gene with a mutated single base, wherein said gene is flanked by an N_4CC nucleotide sequence, wherein said mutated gene causes a genetically-based medical condition; ii) a YE1-BE3-nNme2Cas9 (D16A)-UGI, comprising a mutated NmeCas9 protein, said mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for said N_4CC nucleotide sequence; b) treating said patient with said nNme2Cas9-CBE4 under conditions such that said mutated NmeCas9 protein replaces said mutated single base with a wild type single base, such that said genetically-based medical condition does not develop.

Definitions

To facilitate the understanding of this invention, a number of terms are defined below. Terms defined herein have meanings as commonly understood by a person of ordinary skill in

the areas relevant to the present invention. Terms such as “a”, “an” and “the” are not intended to refer to only a singular entity, but include the general class of which a specific example may be used for illustration. The terminology herein is used to describe specific embodiments of the invention, but their usage does not delimit the invention, except as outlined in the claims.

As used herein, the term “edit” “editing” or “edited” refers to a method of altering a nucleic acid sequence of a polynucleotide (e.g., for example, a wild type naturally occurring nucleic acid sequence or a mutated naturally occurring sequence) by selective deletion of a specific genomic target. Such a specific genomic target includes, but is not limited to, a chromosomal region, a gene, a promoter, an open reading frame or any nucleic acid sequence.

As used herein, the term “single base” refers to one, and only one, nucleotide within a nucleic acid sequence. When used in the context of single base editing, it is meant that the base at a specific position within the nucleic acid sequence is replaced with a different base. This replacement may occur by many mechanisms, including but not limited to, substitution or modification.

As used herein, the term “target” or “target site” refers to a pre-identified nucleic acid sequence of any composition and/or length. Such target sites include, but is not limited to, a chromosomal region, a gene, a promoter, an open reading frame or any nucleic acid sequence. In some embodiments, the present invention interrogates these specific genomic target sequences with complementary sequences of gRNA.

The term “on-target binding sequence” as used herein, refers to a subsequence of a specific genomic target that may be completely complementary to a programmable DNA binding domain and/or a single guide RNA sequence.

The term “off-target binding sequence” as used herein, refers to a subsequence of a specific genomic target that may be partially complementary to a programmable DNA binding domain and/or a single guide RNA sequence.

The term “effective amount” as used herein, refers to a particular amount of a pharmaceutical composition comprising a therapeutic agent that achieves a clinically beneficial result (i.e., for example, a reduction of symptoms). Toxicity and therapeutic efficacy of such compositions can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio

between toxic and therapeutic effects is the therapeutic index, and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and additional animal studies can be used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The term “symptom”, as used herein, refers to any subjective or objective evidence of disease or physical disturbance observed by the patient. For example, subjective evidence is usually based upon patient self-reporting and may include, but is not limited to, pain, headache, visual disturbances, nausea and/or vomiting. Alternatively, objective evidence is usually a result of medical testing including, but not limited to, body temperature, complete blood count, lipid panels, thyroid panels, blood pressure, heart rate, electrocardiogram, tissue and/or body imaging scans.

The term “disease” or “medical condition”, as used herein, refers to any impairment of the normal state of the living animal or plant body or one of its parts that interrupts or modifies the performance of the vital functions. Typically manifested by distinguishing signs and symptoms, it is usually a response to: i) environmental factors (as malnutrition, industrial hazards, or climate); ii) specific infective agents (as worms, bacteria, or viruses); iii) inherent defects of the organism (as genetic anomalies); and/or iv) combinations of these factors.

The terms "reduce," "inhibit," "diminish," "suppress," "decrease," “prevent” and grammatical equivalents (including “lower,” “smaller,” etc.) when in reference to the expression of any symptom in an untreated subject relative to a treated subject, mean that the quantity and/or magnitude of the symptoms in the treated subject is lower than in the untreated subject by any amount that is recognized as clinically relevant by any medically trained personnel. In one embodiment, the quantity and/or magnitude of the symptoms in the treated subject is at least 10% lower than, at least 25% lower than, at least 50% lower than, at least 75% lower than, and/or at least 90% lower than the quantity and/or magnitude of the symptoms in the untreated subject.

The term "attached" as used herein, refers to any interaction between a medium (or carrier) and a drug. Attachment may be reversible or irreversible. Such attachment includes, but

is not limited to, covalent bonding, ionic bonding, Van der Waals forces or friction, and the like. A drug is attached to a medium (or carrier) if it is impregnated, incorporated, coated, in suspension with, in solution with, mixed with, *etc.*

The term "drug" or "compound" as used herein, refers to any pharmacologically active substance capable of being administered which achieves a desired effect. Drugs or compounds can be synthetic or naturally occurring, non-peptide, proteins or peptides, oligonucleotides or nucleotides, polysaccharides or sugars.

The term "administered" or "administering", as used herein, refers to any method of providing a composition to a patient such that the composition has its intended effect on the patient. An exemplary method of administering is by a direct mechanism such as, local tissue administration (*i.e.,* for example, extravascular placement), oral ingestion, transdermal patch, topical, inhalation, suppository *etc*.

The term "patient" or “subject”, as used herein, is a human or animal and need not be hospitalized. For example, out-patients, persons in nursing homes are "patients." A patient may comprise any age of a human or non-human animal and therefore includes both adult and juveniles (*i.e.,* children). It is not intended that the term "patient" connote a need for medical treatment, therefore, a patient may voluntarily or involuntarily be part of experimentation whether clinical or in support of basic science studies.

The term “affinity” as used herein, refers to any attractive force between substances or particles that causes them to enter into and remain in chemical combination. For example, an inhibitor compound that has a high affinity for a receptor will provide greater efficacy in preventing the receptor from interacting with its natural ligands, than an inhibitor with a low affinity.

The term "pharmaceutically" or "pharmacologically acceptable", as used herein, refer to molecular entities and compositions that do not produce adverse, allergic, or other untoward reactions when administered to an animal or a human.

The term, "pharmaceutically acceptable carrier", as used herein, includes any and all solvents, or a dispersion medium including, but not limited to, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils, coatings, isotonic and absorption delaying agents, liposome,

commercially available cleansers, and the like. Supplementary bioactive ingredients also can be incorporated into such carriers.

The term "viral vector" encompasses any nucleic acid construct derived from a virus genome capable of incorporating heterologous nucleic acid sequences for expression in a host organism. For example, such viral vectors may include, but are not limited to, adeno-associated viral vectors, lentiviral vectors, SV40 viral vectors, retroviral vectors, adenoviral vectors. Although viral vectors are occasionally created from pathogenic viruses, they may be modified in such a way as to minimize their overall health risk. This usually involves the deletion of a part of the viral genome involved with viral replication. Such a virus can efficiently infect cells but, once the infection has taken place, the virus may require a helper virus to provide the missing proteins for production of new virions. Preferably, viral vectors should have a minimal effect on the physiology of the cell it infects and exhibit genetically stable properties (e.g., do not undergo spontaneous genome rearrangement). Most viral vectors are engineered to infect as wide a range of cell types as possible. Even so, a viral receptor can be modified to target the virus to a specific kind of cell. Viruses modified in this manner are said to be pseudotyped. Viral vectors are often engineered to incorporate certain genes that help identify which cells took up the viral genes. These genes are called marker genes. For example, a common marker gene confers antibiotic resistance to a certain antibiotic.

As used herein the "*ROSA26* gene" or "*Rosa26* gene" refers to a human or mouse (respectively) locus that is widely used for achieving generalized expression in the mouse. Targeting to the *ROSA26* locus may be achieved by introducing a desired gene into the first intron of the locus, at a unique XbaI site approximately 248 bp upstream of the original gene trap line. A construct may be constructed using an adenovirus splice acceptor followed by a gene of interest and a polyadenylation site inserted at the unique XbaI site. A neomycin resistance cassette may also be included in the targeting vector.

As used herein the "*PCSK9* gene" or "*Pcsk9* gene" refers to a human or mouse (respectively) locus that encodes a PCSK9 protein. The *PCSK9* gene resides on chromosome 1 at the band 1p32.3 and includes 13 exons. This gene may produce at least two isoforms through alternative splicing.

The term "proprotein convertase subtilisin/kexin type 9" and "PCSK9" refers to a protein encoded by a gene that modulates low density lipoprotein levels. Proprotein convertase

subtilisin/kexin type 9, also known as PCSK9, is an enzyme that in humans is encoded by the *PCSK9* gene. Seidah et al., "The secretory proprotein convertase neural apoptosis-regulated convertase 1 (NARC-1): liver regeneration and neuronal differentiation" *Proc. Natl. Acad. Sci. U.S.A.* 100 (3): 928–933 (2003). Similar genes (orthologs) are found across many species. Many enzymes, including PSCK9, are inactive when they are first synthesized, because they have a section of peptide chains that blocks their activity; proprotein convertases remove that section to activate the enzyme. PSCK9 is believed to play a regulatory role in cholesterol homeostasis. For example, PCSK9 can bind to the epidermal growth factor-like repeat A (EGF-A) domain of the low-density lipoprotein receptor (LDL-R) resulting in LDL-R internalization and degradation. Clearly, it would be expected that reduced LDL-R levels result in decreased metabolism of LDL-C, which could lead to hypercholesterolemia.

The term “hypercholesterolemia” as used herein, refers to any medical condition wherein blood cholesterol levels are elevated above the clinically recommended levels. For example, if cholesterol is measured using low density lipoproteins (LDLs), hypercholesterolemia may exist if the measured LDL levels are above, for example, approximately 70 mg/dl. Alternatively, if cholesterol is measured using free plasma cholesterol, hypercholesterolemia may exist if the measured free cholesterol levels are above, for example, approximately 200-220 mg/dl.

As used herein, the term “CRISPRs” or “Clustered Regularly Interspaced Short Palindromic Repeats” refers to an acronym for DNA loci that contain multiple, short, direct repetitions of base sequences. Each repetition contains a series of bases followed by 30 or so base pairs known as "spacer DNA". The spacers are short segments of DNA from a virus and may serve as a 'memory' of past exposures to facilitate an adaptive defense against future invasions.

As used herein, the term “Cas” or “CRISPR-associated (cas)” refers to genes often associated with CRISPR repeat-spacer arrays.

As used herein, the term “Cas9” refers to a nuclease from Type II CRISPR systems, an enzyme specialized for generating double-strand breaks in DNA, with two active cutting sites (the HNH and RuvC domains), one for each strand of the double helix. Jinek combined tracrRNA and spacer RNA into a "single-guide RNA" (sgRNA) molecule that, mixed with Cas9, could find and cleave DNA targets through Watson-Crick pairing between the guide sequence within the sgRNA and the target DNA sequence.

The term “protospacer adjacent motif” (or PAM) as used herein, refers to a DNA sequence that may be required for a Cas9/sgRNA to form an R-loop to interrogate a specific DNA sequence through Watson-Crick pairing of its guide RNA with the genome. The PAM specificity may be a function of the DNA-binding specificity of the Cas9 protein (e.g., a “protospacer adjacent motif recognition domain” at the C-terminus of Cas9).

As used herein, the term “sgRNA” refers to single guide RNA used in conjunction with CRISPR associated systems (Cas). sgRNAs are a fusion of crRNA and tracrRNA and contain nucleotides of sequence complementary to the desired target site. Jinek et al., “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity” *Science* 337(6096):816-821 (2012) Watson-Crick pairing of the sgRNA with the target site permits R-loop formation, which in conjunction with a functional PAM permits DNA cleavage or in the case of nuclease-deficient Cas9 allows binds to the DNA at that locus.

As used herein, the term “fluorescent protein” refers to a protein domain that comprises at least one organic compound moiety that emits fluorescent light in response to the appropriate wavelengths. For example, fluorescent proteins may emit red, blue and/or green light. Such proteins are readily commercially available including, but not limited to: i) mCherry (Clonetech Laboratories): excitation: 556/20 nm (wavelength/bandwidth); emission: 630/91 nm; ii) sfGFP (Invitrogen): excitation: 470/28 nm; emission: 512/23 nm; iii) TagBFP (Evrogen): excitation 387/11 nm; emission 464/23 nm.

As used herein, the term “sgRNA” refers to single guide RNA used in conjunction with CRISPR associated systems (Cas). sgRNAs contains nucleotides of sequence complementary to the desired target site. Watson-crick pairing of the sgRNA with the target site recruits the nuclease-deficient Cas9 to bind the DNA at that locus.

As used herein, the term “orthogonal” refers targets that are non-overlapping, uncorrelated, or independent. For example, if two orthogonal nuclease-deficient Cas9 gene fused to different effector domains were implemented, the sgRNAs coded for each would not cross-talk or overlap. Not all nuclease-deficient Cas9 genes operate the same, which enables the use of orthogonal nuclease-deficient Cas9 gene fused to a different effector domains provided the appropriate orthogonal sgRNAs.

As used herein, the term “phenotypic change” or “phenotype” refers to the composite of an organism's observable characteristics or traits, such as its morphology, development,

biochemical or physiological properties, phenology, behavior, and products of behavior. Phenotypes result from the expression of an organism's genes as well as the influence of environmental factors and the interactions between the two.

"Nucleic acid sequence" and "nucleotide sequence" as used herein refer to an oligonucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand.

The term "an isolated nucleic acid", as used herein, refers to any nucleic acid molecule that has been removed from its natural state (e.g., removed from a cell and is, in a preferred embodiment, free of other genomic nucleic acid).

The terms "amino acid sequence" and "polypeptide sequence" as used herein, are interchangeable and to refer to a sequence of amino acids.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

The term "portion" when used in reference to a nucleotide sequence refers to fragments of that nucleotide sequence. The fragments may range in size from 5 nucleotide residues to the entire nucleotide sequence minus one nucleic acid residue.

As used herein, the terms "complementary" or "complementarity" are used in reference to "polynucleotides" and "oligonucleotides" (which are interchangeable terms that refer to a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "C-A-G-T," is complementary to the sequence "G-T-C-A." Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The terms "homology" and "homologous" as used herein in reference to nucleotide sequences refer to a degree of complementarity with other nucleotide sequences. There may be

partial homology or complete homology (i.e., identity). A nucleotide sequence which is partially complementary, i.e., "substantially homologous," to a nucleic acid sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence. The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

The terms “homology” and “homologous” as used herein in reference to amino acid sequences refer to the degree of identity of the primary structure between two amino acid sequences. Such a degree of identity may be directed to a portion of each amino acid sequence, or to the entire length of the amino acid sequence. Two or more amino acid sequences that are “substantially homologous” may have at least 50% identity, preferably at least 75% identity, more preferably at least 85% identity, most preferably at least 95%, or 100% identity.

An oligonucleotide sequence which is a "homolog" is defined herein as an oligonucleotide sequence which exhibits greater than or equal to 50% identity to a sequence, when sequences having a length of 100 bp or larger are compared.

Low stringency conditions comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5 x SSPE (43.8 g/l NaCl, 6.9 g/l $NaH_2PO_4 \cdot H_2O$ and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5x Denhardt's reagent {50x Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)} and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5x SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed. Numerous equivalent conditions may also be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (

DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol), as well as components of the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, conditions which promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) may also be used.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids using any process by which a strand of nucleic acid joins with a complementary strand through base pairing to form a hybridization complex. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bounds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., C_0 t or R_0 t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized to a solid support (e.g., a nylon membrane or a nitrocellulose filter as employed in Southern and Northern blotting, dot blotting or a glass slide as employed in in situ hybridization, including FISH (fluorescent in situ hybridization)).

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring. An end of an oligonucleotide is referred to as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of another mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA

molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements which direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

The term "transfection" or "transfected" refers to the introduction of foreign DNA into a cell.

As used herein, the terms "nucleic acid molecule encoding", "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

As used herein, the term "gene" means the deoxyribonucleotide sequences comprising the coding region of a structural gene and including sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene which are transcribed into heterogeneous nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking

region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

The term "label" or "detectable label" are used herein, to refer to any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Such labels include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads®), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and calorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include, but are not limited to, U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241 (all herein incorporated by reference in their entirety). The labels contemplated in the present invention may be detected by many methods. For example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting, the reaction product produced by the action of the enzyme on the substrate, and calorimetric labels are detected by simply visualizing the colored label.

Brief Description Of The Drawings

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

Figure 1 illustrates exemplary schematic embodiments of an NmeCas9 deaminase fusion protein single base editor and exemplary constructed plasmids of base editors.

Figure 1A shows an exemplary YE1-BE3-nNme2Cas9 (D16A)-UGI construct.

Figure 1B shows an exemplary ABE7.10 nNme2Cas9 (D16A) construct.

Figure 1C shows an exemplary ABE7.10-nNme2Cas9 (D16A) construct comprising two SV40 NLS sequences.

Figure 1D shows an exemplary nNme2Cas9-CBE4 (also called a BE4-nNme2Cas9 (D16A)-UGI-UGI) construct.

Figure 1E shows an exemplary optimized nNme2Cas9-ABEmax construct.

Figure 2 presents exemplary data of the electroporation of HEK293T cells with DNA plasmids comprising a YE1-BE3-nNme2Cas9 (D16A)-UGI fusion protein efficiently converting C to T at endogenous target site 25 (TS25) in HEK293T cells via nucleofection.

Figure 2A shows exemplary sequences for a TS25 endogenous target site (within the black rectangle). GN23 sgRNA base-pairs with the target DNA strand, leaving the displaced DNA strand for cytidine deaminase to edit (e.g. new green nucleotides).

Figure 2B shows exemplary sequencing data showing a doublet nucleotide peak (7th position from 5' end; arrow) demonstrating the successful single base editing of a cytidine to a thymidine (e.g., a C•G base pair conversion to a T•A base pair).

Figure 2C shows an exemplary quantitation of the data shown in Figure 2B plotting the percent conversion of C → T single base editing. The percentage of C converted to T is about 40% in the base editor- and sgRNA-treated sample (p-value = 6.88 x 10-6). The "no sgRNA" control displays the background noise due to Sanger sequencing. EditR (Kluesner et al., 2018) was used to perform the analysis.

Figure 3 presents exemplary specific UGI target sites that were respectively integrated into YE1-BE3-nNme2Cas9/D16A mutant fusion proteins and co-expressed with enhanced green fluorescent protein (EGFP) in a stable K562-derived cell line. Converted bases are highlighted in orange color. Background signals were filtered using negative control samples (YE1-BE3-nNme2Cas9 nucleofected K562 cells without sgRNA constructs). N_4CC PAMs are boxed. The percentage of total reads exhibiting mutations in base-editor-targeted sites is shown in the right column.

Figure 3A shows an exemplary EGFP-Site 1.

Figure 3B shows an exemplary EGFP-Site 2.

Figure 3C shows an exemplary EGFP-Site 3.

Figure 3D shows an exemplary EGFP-Site 4.

Figure 3E shows an exemplary deep-sequencing analysis indicating where YE1-BE3-nNme2Cas9 converts C residues to T residues at endogenous *c-fos* promoter region. The percentage of total reads exhibiting mutations in base-editor-targeted

sites is shown in the right column. The converted bases are highlighted in orange or yellow color. Background signals were filtered using negative control samples. The highest percentage of editing is 32.50%.

Figure 3F shows an exemplary deep-sequencing analysis indicating where ABE7.10-nNme2Cas9 or ABEmax (Koblan *et al.,* 2018)-nNme2Cas9 converts A residues to G residues at endogenous *c-fos* promoter region. The percentage of total reads exhibiting mutations in base-editor-targeted sites is shown in the right column. The converted bases are highlighted in orange color. Background signals were filtered using negative control samples. The percentage of editing is 0.53% by ABE7.10-nNme2Cas9 or 2.33% by ABEmax-nNme2Cas9 (D16A).

Figure 4 presents an exemplary alignment of the wildtype *Fah* gene with the tyrosinemia *Fah* mutant gene showing an A-G single base gene editing target site (position 9). The respective SpyCas9 single PAM site and NmeCas9 double PAM sites are indicated for demonstrating the suboptimal targeting window relative to the SpyCas9 PAM site.

Figure 5 illustrates exemplary three closely related *Neisseria meningitidis* Cas9 orthologs that have distinct PAMs.

Figure 5A shows an exemplary schematic showing mutated residues (orange spheres) between Nme2Cas9 (left) and Nme3Cas9 (right) mapped onto the predicted structure of Nme1Cas9, revealing the cluster of mutations in the PID (black).

Figure 5B shows an exemplary experimental workflow of the in vitro PAM discovery assay with a 10-bp randomized PAM region. Following in vitro digestion, adapters were ligated to cleaved products for library construction and sequencing.

Figure 5C shows exemplary sequence logos resulting from in vitro PAM discovery reveal the enrichment of a N_4GATT PAM for Nme1Cas9, consistent with its previously established specificity.

Figure 5D shows exemplary sequence logos indicating that Nme1Cas9 with its PID swapped with that of Nme2Cas9 (left) or Nme3Cas9 (right) requires a C at PAM position 5. The remaining nucleotides were not determined with high confidence due to the modest cleavage efficiency of the PID-swapped protein chimeras (see Figure 6C).

Figure 5E shows an exemplary sequence logo showing that full-length Nme2Cas9 recognizes an N_4CC PAM, based on efficient substrate cleavage of a target pool with a fixed C at PAM position 5, and with PAM nts 1-4 and 6-8 randomized.

Figure 6 presents a characterization of *Neisseria meningitidis* Cas9 orthologs with rapidly-evolving PIDs, as related to Figure 5**.**

Figure 6A shows an exemplary unrooted phylogenetic tree of NmeCas9 orthologs that are >80% identical to Nme1Cas9. Three distinct branches emerged, with the majority of mutations clustered in the PID. Groups 1 (blue), 2 (orange), and 3 (green) have PIDs with >98%, ~52%, and ~86% identity to Nme1Cas9, respectively. Three representative Cas9 orthologs (one from each group) (Nme1Cas9, Nme2Cas9 and Nme3Cas9) are indicated.

Figure 6B shows an exemplary schematic showing the CRISPR-*cas* loci of the strains encoding the three Cas9 orthologs (Nme1Cas9, Nme2Cas9, and Nme3Cas9) from (A). Percent identities of each CRISPR-Cas component with *N. meningitidis* 8013 (encoding Nme1Cas9) are shown. Blue and red arrows denote pre-crRNA and tracrRNA transcription initiation sites, respectively.

Figure 6C shows an exemplary normalized read counts (% of total reads) from cleaved DNAs from the in vitro assays for intact Nme1Cas9 (grey), for chimeras with Nme1Cas9's PID swapped with those of Nme2Cas9 and Nme3Cas9 (mixed colors), and for full-length Nme2Cas9 (orange), are plotted. The reduced normalized read counts indicate lower cleavage efficiencies in the chimeras.

Figure 6D shows an exemplary sequence logos from the in vitro PAM discovery assay on an NNNNCNNN PAM pool by Nme1Cas9 with its PID swapped with those of Nme2Cas9 (left) or Nme3Cas9 (right).

Figure 7 presents exemplary data showing that Nme2Cas9 uses a 22-24 nt spacer to edit sites adjacent to an N_4CC PAM. All experiments were done in triplicate, and error bars represent the standard error of the mean (s.e.m.).

Figure 7A shows an exemplary schematic diagram depicting transient transfection and editing of HEK293T TLR2.0 cells, with mCherry+ cells detected by flow cytometry 72 hours after transfection.

Figure 7B shows an exemplary Nme2Cas9 editing of the TLR2.0 reporter. Sites with N_4CC PAMs were targeted with varying efficiencies, while no Nme2Cas9 targeting was observed at an N_4GATT PAM or in the absence of sgRNA. SpyCas9 (targeting a previously validated site with an NGG PAM) and Nme1Cas9 (targeting N_4GATT) were used as positive controls.

Figure 7C shows an exemplary effect of spacer length on the efficiency of Nme2Cas9 editing. An sgRNA targeting a single TLR2.0 site, with spacer lengths varying from 24 to 20 nts (including the 5'-terminal G required by the U6 promoter), indicate that highest editing efficiencies are obtained with 22-24 nt spacers.

Figure 7D shows an exemplary An Nme2Cas9 dual nickase can be used in tandem to generate NHEJ- and HDR-based edits in TLR2.0. Nme2Cas9- and sgRNA-expressing plasmids, along with an 800-bp dsDNA donor for homologous repair, were electroporated into HEK293T TLR2.0 cells, and both NHEJ (mCherry+) and HDR (GFP+) outcomes were scored by flow cytometry. HNH nickase, Nme2Cas9^{D16A}; RuvC nickase, Nme2Cas9^{H588A}. Cleavage sites 32 bp and 64 bp apart were targeted using either nickase. The HNH nickase (Nme2Cas9^{D16A}) yielded efficient editing, particularly with the cleavage sites that were separated by 32 bp, whereas the RuvC nickase (Nme2Cas9^{H588A}) was not effective. Wildtype Nme2Cas9 was used as a control.

Figure 8 presents exemplary data showing PAM, spacer, and seed requirements for Nme2Cas9 targeting in mammalian cells, as related to Figure 7. All experiments were done in triplicate and error bars represent s.e.m.

Figure 8A shows an exemplary Nme2Cas9 targeting at N_4C<u>D</u> sites in TLR2.0, with editing estimated based on mCherry+ cells. Four sites for each non-C nucleotide at the tested position (N_4C<u>A</u>, N_4C<u>T</u> and N_4C<u>G</u>) were examined, and an N_4CC site was used as a positive control.

Figure 8B shows an exemplary Nme2Cas9 targeting at N_4<u>D</u>C sites in TLR2.0 [similar to (A)].

Figure 8C shows exemplary guide truncations on a TLR2.0 site (distinct from that in Figure 2C) with a N_4CCA PAM, revealing similar length requirements as those observed at the other site.

Figure 8D shows exemplary Nme2Cas9 targeting efficiency is differentially sensitive to single-nucleotide mismatches in the seed region of the sgRNA. Data show the effects of walking single-nucleotide sgRNA mismatches along the 23-nt spacer in a TLR2.0 target site.

Figure 9 presents exemplary data showing Nme2Cas9 genome editing at endogenous loci in mammalian cells via multiple delivery methods. All results represent 3 independent biological replicates, and error bars represent s.e.m.

Figure 9A shows an exemplary Nme2Cas9 genome editing of endogenous human sites in HEK293T cells following transient transfection of Nme2Cas9- and sgRNA-expressing plasmids. 40 sites were screened initially (Table 1); the 14 sites shown (selected to include representatives of varying editing efficiencies, as measured by TIDE) were then re-analyzed in triplicate. An Nme1Cas9 target site (with an N_4GATT PAM) was used as a negative control.

Figure 9B shows exemplary data charts: Left panel: Transient transfection of a single plasmid expressing both Nme2Cas9 and sgRNA (targeting the *Pcsk9* and *Rosa26* loci) enables editing in Hepa1-6 mouse cells, as detected by TIDE. Right panel: Electroporation of sgRNA plasmids into K562 cells stably expressing Nme2Cas9 from a lentivector results in efficient indel formation.

Figure 9C shows exemplary Nme2Cas9 can be electroporated as an RNP complex to induce genome editing. 40 picomoles Cas9 along with 50 picomoles of in vitro-transcribed sgRNAs targeting three different loci were electroporated into HEK293T cells. Indels were measured after 72h using TIDE.

Figure 10 presents exemplary data showing dose dependence and segmental deletions by Nme2Cas9, as related to Figure 9.

Figure 10A shows exemplary increasing the dose of electroporated Nme2Cas9 plasmid (500 ng, vs. 200 ng in Figure 3A) improves editing efficiency at two sites (TS16 and TS6). Data provided in yellow are re-used from Figure 9A.

Figure 10B shows exemplary Nme2Cas9 can be used to create precise segmental deletions. Two TLR2.0 targets with cleavage sites 32 bp apart were targeted simultaneously with Nme2Cas9. The majority of lesions created were deletions of exactly 32 bp (blue).

Figure 11 presents exemplary data showing that Nme2Cas9 is subject to inhibition by a subset of type II-C anti-CRISPR families in vitro and in cells. All experiments were done in triplicate and error bars represent s.e.m.

Figure 11A shows exemplary In vitro cleavage assay of Nme1Cas9 and Nme2Cas9 in the presence of five previously characterized anti-CRISPR proteins (10:1 ratio of Acr:Cas9). Top: Nme1Cas9 efficiently cleaves a fragment containing a protospacer with an N_4GATT PAM in the absence of an Acr or in the presence of a negative control Acr (AcrE2). All five previously characterized type II-C Acr families inhibited Nme1Cas9, as expected. Bottom: Nme2Cas9 inhibition mirrors that of Nme1Cas9, except for the lack of inhibition by AcrIIC5$_{Smu}$.

Figure 11B shows exemplary genome editing in the presence of the five previously described anti-CRISPR families. Plasmids expressing Nme2Cas9 (200 ng), sgRNA (100 ng) and each respective Acr (200 ng) were co-transfected into HEK293T cells, and genome editing was measured using Tracking of Indels by Decompostion (TIDE) 72 hr post transfection. Consistent with our in vitro analyses, all type II-C anti-CRISPRs except AcrIIC5$_{Smu}$ inhibited genome editing, albeit with different efficiencies.

Figure 11C shows exemplary Acr inhibition of Nme2Cas9 is dose-dependent with distinct apparent potencies. Nme2Cas9 is fully inhibited by AcrIIC1$_{Nme}$ and AcrIIC4$_{Hpa}$ at 2:1 and 1:1 mass ratios of cotransfected Acr and Nme2Cas9 plasmids, respectively.

Figure 12 presents exemplary data showing that a Nme2Cas9 PID swap renders Nme1Cas9 insensitive to AcrIIC5$_{Smu}$ inhibition, as related to Figure 11. In vitro cleavage by the Nme1Cas9-Nme2Cas9PID chimera in the presence of previously characterized Acr proteins (10 uM Cas9-sgRNA + 100 uM Acr).

Figure 13 presents exemplary data showing orthogonality and relative accuracy of Nme2Cas9 and SpyCas9 at dual target sites, as related to Figure 12.

Figure 13A shows exemplary Nme2Cas9 and SpyCas9 guides are orthogonal. TIDE results show the frequencies of indels created by both nucleases targeting DS2 with either their cognate sgRNAs or with the sgRNAs of the other ortholog.

Figure 13B shows exemplary Nme2Cas9 and SpyCas9 exhibiting comparable on-target editing efficiencies as assessed by GUIDE-seq. Bars indicate on-target read counts from GUIDE-Seq at the three dual sites targeted by each ortholog. Orange bars represent Nme2Cas9 and black bars represent SpyCas9.

Figure 13C shows an exemplary SpyCas9's on-target vs. off-target read counts for each site. Orange bars represent the on-target reads while black bars represent off-targets.

Figure 13D shows exemplary Nme2Cas9's on-target vs. off-target reads for each site.

Figure 13E bar graphs showing exemplary indel efficiencies (measured by TIDE) at potential off-target sites predicted by CRISPRSeek. On- and off-target site sequences are shown on the left, with the PAM region underlined and sgRNA mismatches and non-consensus PAM nucleotides given in red.

Figure 14 presents exemplary data showing that Nme2Cas9 exhibits little or no detectable off-targeting in mammalian cells.

Figure 14Ashows an exemplary schematic depicting dual sites (DSs) targetable by both SpyCas9 and Nme2Cas9 by virtue of their non-overlapping PAMs. The Nme2Cas9 PAM (orange) and SpyCas9 PAM (blue) are highlighted. A 24nt Nme2Cas9 guide sequence is indicated in yellow; the corresponding guide sequence for SpyCas9 would be 4nt shorter at the 5' end.

Figure 14B shows an exemplary Nme2Cas9 and SpyCas9 that both induce indels at DSs. Six DSs in *VEGFA* (with $GN_3GN_{19}NGGNCC$ sequences) were selected for direct comparisons of editing by the two orthologs. Plasmids expressing each Cas9 (with the same promoter, linkers, tags and NLSs) and its cognate guide were transfected into HEK293T cells. Indel efficiencies were determined by TIDE 72 hrs post transfection. Nme2Cas9 editing was detectable at all six sites and was marginally or significantly more efficient than SpyCas9 at two sites (DS2 and DS6, respectively). SpyCas9 edited four out of the six sites (DS1, DS2, DS4 and DS6), with two sites showing significantly higher editing efficiencies than Nme2Cas9 (DS1 and DS4). DS2, DS4 and DS6 were selected for GUIDE-Seq analysis as Nme2Cas9 was equally efficient, less efficient and more efficient than SpyCas9, respectively, at these sites.

Figure 14C shows exemplary Nme2Cas9 genome editing that is highly accurate in human cells. Numbers of off-target sites detected by GUIDE-Seq for each nuclease at individual target sites are shown. In addition to dual sites, we analyzed TS6 (because of its high on-target editing efficiency) and *Pcsk9* and *Rosa26* sites in mouse Hepa1-6 cells (to measure accuracy in another cell type).

Figure 14D shows an exemplary targeted deep sequencing to detect indels in edited cells confirms the high Nme2Cas9 accuracy indicated by GUIDE-seq.

Figure 14E shows an exemplary sequence for the validated off-target site of the *Rosa26* guide, showing the PAM region (underlined), the consensus CC PAM dinucleotide (bold), and three mismatches in the PAM-distal portion of the spacer (red).

Figure 15 presents exemplary data showing Nme2Cas9 genome editing in vivo via all-in-one AAV delivery.

Figure 15A shows exemplary workflow for delivery of AAV8.sgRNA.Nme2Cas9 to lower cholesterol levels in mice by targeting *Pcsk9*. Top: schematic of the all-in-one AAV vector expressing Nme2Cas9 and the sgRNA (individual genome elements not to scale). BGH, bovine growth hormone poly(A) site; HA, epitope tag; NLS, nuclear localization sequence; h, human-codon-optimized. Bottom: Timeline for AAV8.sgRNA.Nme2Cas9 tail-vein injections (4×10^{11} GCs), followed by cholesterol measurements at day 14 and indel, histology and cholesterol analyses at day 28 post-injection.

Figure 15B shows an exemplary TIDE analysis to measure indels in DNA extracted from livers of mice injected with AAV8.Nme2Cas9+sgRNA targeting *Pcsk9* and *Rosa26* (control) loci. Indel efficiency at the lone off-target site identified by GUIDE-seq for these two sgRNAs (Rosa26|OT1) were also assessed by TIDE.

Figure 15C shows an exemplary reduced serum cholesterol levels in mice injected with the *Pcsk9*-targeting guide compared to the *Rosa26*-targeting controls. *P* values are calculated by unpaired two-tailed t-test.

Figure 16 presents exemplary data showing PCSK9 knockdown and liver histology following Nme2Cas9 AAV delivery and editing, related to Figure 15.

Figure 16A shows exemplary Western blotting using anti-PCSK9 antibody reveals strongly reduced levels of PCSK9 in the livers of mice treated with sg*Pcsk9*, compared to mice treated with sg*Rosa26.* 2ng of recombinant PCSK9 was used as a mobility standard (left-most lane), and a cross-reacting band in the liver samples is indicated by an asterisk. GAPDH was used as loading control (bottom panel).

Figure 16B shows exemplary H&E staining from livers of mice injected with AAV8.Nme2Cas9+sg*Rosa26* (left) or AAV8.Nme2Cas9+sg*Pcsk9* (right) vectors. Scale bars, 25 μm.

Figure 17 presents exemplary data showing *Tyr* editing ex vivo in mouse zygotes, related to Figure 16.

Figure 17A shows an exemplary two sites in *Tyr*, each with N_4CC PAMs, were tested for editing in Hepa1-6 cells. The sg*Tyr2* guide exhibited higher editing efficiency and was selected for further testing.

Figure 17B shows an exemplary seven mice that survived post-natal development, and each exhibited coat color phenotypes as well as on-target editing, as assayed by TIDE.

Figure 17C shows an exemplary Indel spectra from tail DNA of each mouse from (B), as well as an unedited C57BL/6NJ mouse, as indicated by TIDE analysis. Efficiencies of insertions (positive) and deletions (negative) of various sizes are indicated.

Figure 18 presents exemplary data showing Nme2Cas9 genome editing ex vivo via all-in-one AAV delivery.

Figure 18A shows an exemplary workflow for single-AAV Nme2Cas9 editing ex vivo to generate albino C57BL/6NJ mice by targeting the *Tyr* gene. Zygotes are cultured in KSOM containing AAV6.Nme2Cas9:sg*Tyr* for 5-6 hours, rinsed in M2, and cultured for a day before being transferred to the oviduct of pseudo-pregnant recipients.

Figure 18B shows exemplary albino (left) and chinchilla or variegated (middle) mice generated by $3x10^9$ GCs, and chinchilla or variegated mice (right) generated by $3x10^8$ GCs of zygotes with AAV6.Nme2Cas9:sg*Tyr*.

Figure 18C shows an exemplary summary of Nme2Cas9.sg*Tyr* single-AAV ex vivo *Tyr* editing experiments at two AAV doses.

Figure 19 shows an exemplary mCherry reporter assay for nSpCas9-ABEmax and optimized ABEmax-nNme2Cas9 (D16A) activities.

Figure 19A shows exemplary sequence information of sequence information of ABE-mCherry reporter. There is a TAG stop codon in mCherry coding region. In the reporter-integrated stable cell line, there is no mCherry signal. The mCherry signal will show up if the nSpCas9-ABEmax or optimized **ABEmax-nNme2Cas9 (D16A)** can convert TAG to CAG (which is encoded Gln).

Figure 19B shows an exemplary mCherry signals light up since SpCas9-ABE or **ABEmax-nNme2Cas9 (D16A)** is active in the specific region of the mCherry reporter. Upper panel is the negative control, middle panel shows the mCherry signals light up in reporter cells treated with nSpCas9-ABEmax, bottom panel shows the mCherry signals light up in reporter cells treated with optimized **ABEmax-nNme2Cas9 (D16A)**.

Figure 19C shows an exemplary FACs Quantitation of base editing events in mCherry reporter cells transfected with the SpCas9-ABE or **ABEmax-nNme2Cas9 (D16A)**. N = 6; error bars represent S.D. Results are from biological replicates performed in technical duplicates.

Figure 20 shows an exemplary GFP reporter assay for nSpCas9-CBE4 (Addgene #100802) and **CBE4-nNme2Cas9 (D16A)-UGI-UGI** (CBE4 was cloned from Addgene #100802) activities.

Figure 20A shows exemplary sequence information of CBE-GFP reporter. There is a mutation in the fluorophore core region of the GFP reporter line, which converts GYG to GHG. Therefore, there is no GFP signal. The GFP signal will show up if the nSpCas9-CBE4 or **CBE4-nNme2Cas9 (D16A)-UGI-UGI** can convert CAC to TAC/TAT (Histidine to Tyrosine).

Figure 20B shows an exemplary GFP signal (green) since nSpCas9-CBE4 or **CBE4-nNme2Cas9 (D16A)-UGI-UGI** is active in the specific region of the GFP reporter. Upper panel is the negative control. Middle panel shows that the mCherry signals light up in the reporter cells treated with **CBE4-nNme2Cas9 (D16A)-UGI-UGI**. Bottom panel shows that the GFP signals light up in the reporter cells treated with **CBE4-nNme2Cas9 (D16A)-UGI-UGI**).

Figure 20C shows an exemplary FACs Quantitation of base editing events in GFP reporter cells transfected with nSpCas9-CBE4 or **CBE4-nNme2Cas9 (D16A)-UGI-UGI**. N = 6; error bars represent S.D. Results are from biological replicates performed in technical duplicates.

Figure 21 shows exemplary cytosine editing by **CBE4-nNme2Cas9 (D16A)-UGI-UGI**. Upper panel shows the KANK3 targeting sequence information (PAM sequences are indicated in red) of Nme2Cas9 and base editing in the negative control samples. Bottom panel shows the quantification of the substitution rate of each type of base in the **CBE4-nNme2Cas9 (D16A)-UGI-UGI** editing window of the KANK3 target sequences. Sequence tables show nucleotide frequencies at each position. Frequencies of expected C-to-T conversion are highlighted in red.

Figure 22 shows exemplary cytosine and adenine editing by **CBE4-nNme2Cas9 (D16A)-UGI-UGI** and optimized **ABEmax-nNme2Cas9 (D16A)**, respectively. Upper panel shows the PLXNB2 targeting sequence information (PAM sequences are indicated in red) of Nme2Cas9 and base editing in the negative control samples. Middle panel shows the quantification of the substitution rate of each type of base in the optimized **ABEmax-nNme2Cas9 (D16A)** editing windows of the PLXNB2 target sequences. Sequence tables show nucleotide frequencies at each position. Frequencies of expected A-to-G conversion are highlighted in red. Bottom panel shows the quantification of the substitution rate of each type of base in the **CBE4-nNme2Cas9 (D16A)-UGI-UGI** editing windows of the PLXNB2 target sequences. Sequence tables show nucleotide frequencies at each position. Frequencies of expected C-to-T conversion are highlighted in red.

Detailed Description Of The Invention

The present invention is related to the field of gene editing. In particular, the gene editing is directed toward single nucleotide base editing. For example, such single nucleotide base editing results in a conversion of a C•G base pair to a T•A base pair. The high accuracy and precision of the presently disclosed single nucleotide base gene editor is accomplished by an NmeCas9 nuclease that is fused to a nucleotide deaminase protein. The compact nature of the NmeCas9 coupled with a larger number of compatible protospacer adjacent motifs provide the Cas9 fusion constructs contemplated herein can edit sites that are not targetable by conventional SpyCas9 base editor platforms.

A. NmeCas9 Single Base Editing

Cas9 is a programmable nuclease that uses a guide RNA to create a double-stranded break at any desired genomic locus. This programmability has been harnessed for biomedical and therapeutic approaches. However, Cas9-induced breaks often lead to imprecise repair by the cellular machinery, hindering its therapeutic application for single-base corrections as well as uniform and precise gene knock-outs. Moreover, it is extremely challenging to combine Cas9-induced DNA double strand breaks and a repair template for homologous directed repair (HDR) for correcting genetic mutations in post-mitotic cells (e.g. neuronal cells).

Single nucleotide base editing is a genome editing approach where a nuclease-dead or -impaired Cas9 (e.g., dead Cas9 (dCas9) or nickase Cas9 (nCas9)) is fused to another enzyme capable of base-editing nucleotides without causing DNA double strand breaks. To date, two broad classes of Cas9 base editors have been developed: i) cytidine deaminase (edits a C•G base pair to a T•A base pair) SpyCas9 fusion protein; and ii) adenosine deaminase (edits a A•T base pair to a G•C base pair) SpyCas9. Liu et al., "Nucleobase editors and uses thereof" US 2017/0121693; and Lui et al., "Fusions of cas9 domains and nucleic acid-editing domains" US 2015/0166980; (both herein incorporated by reference).

However as mentioned above, SpyCas9 base editing platforms cannot be used to target all single-base mutations due to their limited editing windows. The editing window is constrained by the requirement for an NGG PAM. SpyCas9 is also intrinsically associated with high off-targeting effects in genome editing.

In one embodiment, the present invention contemplates a deaminase fusion protein with a compact and hyper-accurate Nme2Cas9 (*Neisseria meningitidi*s spp.). This Nme2Cas9 has 1,082 amino acids as compared to SpyCas9 that has 1,368 amino acids. This Nme2Cas9 ortholog functions efficiently in mammalian cells, recognizes an N_4CC PAM, and is intrinsically hyper-accurate. Edraki et al., Mol Cell. (in preparation).

Although it is not necessary to understand the mechanism of an invention, it is believed that the compactness and hyper-accuracy of an NmeCas9 base editor targets single-base mutations that could not be reached previously by other Cas9 platforms currently known in the art. It is further believed that the NmeCas9 base editors contemplated herein target pathogenic mutations that are not feasible via current base editor platforms, and with an increased base editing accuracy.

In one embodiment, the present invention contemplates a fusion protein comprising a Nme2Cas9 and a deaminase protein, exemplary examples including ABE7.10-nNme2Cas9 (D16A); Optimized nNme2Cas9-ABEmax; nNme2Cas9-CBE4 (equals BE4-nNme2Cas9 (D16A)-UGI-UGI) as well as ABEmax-nNme2Cas9 (D16A). See, Figure 1A, Figure 1B, Figure 1C, Figure 1D and Figure 1E.

Figure 1 illustrates exemplary schematic embodiments of an NmeCas9 deaminase fusion protein single base editor and exemplary constructed plasmids of base editors. Figure 1A shows an exemplary YE1-BE3-nNme2Cas9 (D16A)-UGI construct. Figure 1B shows an exemplary ABE7.10 nNme2Cas9 (D16A) construct. Figure 1C shows an exemplary ABE7.10-nNme2Cas9 (D16A) construct. Figure 1C shows an exemplary ABE7.10-nNme2Cas9 (D16A) construct comprising two SV40 NLS sequences. Figure 1D shows an exemplary nNme2Cas9-CBE4 (also called a BE4-nNme2Cas9 (D16A)-UGI-UGI) construct. Figure 1E shows an exemplary optimized nNme2Cas9-ABEmax construct.

In one embodiment, the deaminase protein is Apobec1 (YE1-BE3). It is not intended to limit Apobec1 to one organism. In one embodiment, the Apobec1 is derived from a rat species. Kim et al., "Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions". *Nature Biotechnology* 35 (2017). In one embodiment, the Nme2Cas9 comprises an nNme2Cas9 D16A mutant. In one embodiment, the fusion protein further comprises a uracil glycosylase inhibitor protein (UGI). In one embodiment, the fusion protein comprises a YE1-BE3-nNme2Cas9 (D16A)-UGI construct. In one embodiment, the YE1-BE3-nNme2Cas9 (D16A)-UGI construct has the sequence of:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSYSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPENRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTSESATP**
ES*MAAFKPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMAR*
RLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPL
EWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEK
ESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA
VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKS
KLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSS

ELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKR
YDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIET
AREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCL
YSGKEINLVRLNEKGYVEIDAALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNS
REWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKG
KRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAF
DGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLS
SRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMV
NYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKK
NAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSL
HKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNEL
*GKEIRPCRLKKRPPVRSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDIL**
VHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV*

YE1-BE3 (underlined); linker (bold), nNme2Cas9 (italics), UGI (bold/underlined), SV40 NLS (plain).

In one embodiment, the YE1-BE3-nNme2Cas9 (D16A)-UGI construct has the sequence of:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNK
HVEVNFIEKFTTERYFCPNTRCSITWFLSYSPCGECSRAITEFLSRYPHVTLFIYIARLYHH
ADPENRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLE
LYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK**SGSETPGTSESATP**
ES*MAAFKPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMAR*
RLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPL
EWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEK
ESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA
VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKS
KLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSS
ELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKR
YDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIET
AREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCL

YSGKEINLVRLNEKGYVEIDAALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNS
REWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKG
KRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAF
DGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLS
SRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMV
NYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKK
NAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSL
HKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNEL
*GKEIRPCRLKKRPPVR***SGGS**<u>**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDI**</u>
<u>**LVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML**</u>**SGGS**PKKKRKV*

YE1-BE3 (underlined); linker (bold), nNme2Cas9 (italics), UGI (bold/underlined), SV40 NLS (plain).

In one embodiment, the present invention contemplates a fusion protein comprising an NmeCas9/ABE7.10 deaminase protein. In one embodiment, the deaminase protein is TadA. In one embodiment, the deaminase protein is TadA 7.10. In one embodiment, the ABE7.10-nNme2Cas9 (D16A) construct has the following sequence:

<u>MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH</u>
<u>AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG</u>
<u>SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD</u>***SGGSSGG***
SSGSETPGTSESATPESSGGSSGGS**<u>SEVEFSHEYWMRHALTLAKRARDEREVPVGAV</u>**
<u>LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCV</u>
<u>MCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAA</u>
<u>LLCYFFRMPRQVFNAQKKAQSSTDS</u>*GGSSGGSSGSETPGTSESATPESSGGSSGGSMAAF*
KPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVR
RLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLL
HLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRN
QRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLG
HCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQA
RKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIG
TAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEI

YGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSF
KDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLV
RLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKA
RVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASN
GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDK
ETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVH
EYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGR
EIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIA
DNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLI
AFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNELGKEIRP
*CRLKKRPPVR***ED**KRPAATKKAGQAKKKK*

TadA (underlined), TadA 7.10 (underlined/bold), *linker (bold),* nNme2Cas9 (italics), Nucleoplasmin NLS (plain).

In one embodiment, an ABE7.10-nNme2Cas9 (D16A) construct has the following amino acid sequence:

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAG
SLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD***SGGSSGG***
SSGSETPGTSESATPESSGGSSGGS**SEVEFSHEYWMRHALTLAKRARDEREVPVGAV**
LVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCV
MCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAA
LLCYFFRMPRQVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESSGGSSGGS****MA*
AFKPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARS
VRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAV
LLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHI
RNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKM
LGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYA
QARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQD
EIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEAC
AEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVG

KSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEI
NLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQE
FKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVF
ASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTI
DKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEA
VHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKN
GREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYT
IADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSLHKYD
LIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNELGKEIR
*PCRLKKRPPVR**ED***KRPAATKKAGQAKKKK*

TadA (underlined), TadA 7.10 (underlined/bold), *linker (bold italics),* nNme2Cas9 (*italics*), Nucleoplasmin NLS (plain).

In one embodiment, an ABEmax-nNme2Cas9 (D16A) construct has the following amino acid sequence:

***MKRTADGSEFES*PKKKRKV**SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNN
RVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIH
SRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQ
EIKAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESSGGSSGGS***SEVEFSHEYWMRHA
LTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLV
MQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYP
GMNHRVEITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETP*
GTSESATPESSGGSSGGS*MAAFKPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVF*
ERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNT
PWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQT
GDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKE
GIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERP
LTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRA
LEKEGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQIS
LKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVI
NGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSK

DILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNK
GNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRF
LCQFVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQ
QKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEE
ADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVK
RVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKA
VRVEKTQESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDI
DCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQF
*RISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR**ED**KRPAATKKAGQAKKKK**FE**PKKKRK*
V*

TadA (underlined), TadA* 7.10 (underlined/bold), *linker (bold italics),* nNme2Cas9 (italics), Nucleoplasmin NLS (plain) and SV40 NLS (BOLD).

In one embodiment, a **CBE4-nNme2Cas9** (D16A)-UGI-UGI construct has the following amino acid sequence:

PAAKRVKLD***GGSGGGSGGGSG***PAAKRVKLD***GGSGGGSGGGSGPLE*****PKKKRKV*****PW***SSE
TGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEV
NFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPR
NRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCII
LGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK***SGGSSGGSSGSETPGTSE***
SATPESSGGSSGGSIDKL*AAFKPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFER*
AEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPW
QLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGD
FRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIE
TLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTD
TERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEK
EGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKA
LRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGV
VRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILK
LRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQ
TPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQ

FVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKI
TRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADT
PEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVW
LTEIKLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRV
EKTQESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCK
GYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRIST
*QNLVLIQKYQVNELGKEIRPCRLKKRPPVR**VYPYDVPDYAGYPYDVPDYAGSYPYDVPDYA***
***GSAAPAAKKKKLDFESGEFLQPGIDLSQLGGDSGGSGGSGGS*<u>TNLSDIIEKETGKQLVI</u>**
<u>QESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS</u>
<u>NGENKIKML</u>*<u>SGGSGGSGGS</u>*<u>TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD</u>
<u>ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML</u>*<u>SGGSPKKKRKVSR</u>*
<u>GSAA</u>*PAAKRVKLDGGSGGGSGGGSGSG***PAAKRVKLD

rApobec1 (underlined), UGI (underlined/bold), linker *(bold italics),* nNme2Cas9 (D16A) (*italics*), Cmyc-NLS (plain) and SV40 NLS (BOLD).

In one embodiment, an optimized nNme2Cas9-ABEmax construct refers to an optimized version with improved promoter, NLS sequences, and linker sequences. In some embodiments, an optimized nNme2Cas9-ABEmax construct comprises, 5' to 3', a C-myc NLS, 12aa linker, 15aa linker, SV40 NLS, TadA, TadA*7.10, 48aa linker, nNme2Cas9, a 73aa linker (3xHA-tag), 15aa linker, and a C-myc NLS. In some embodiments, an optimized nNme2Cas9-ABEmax construct further comprises at least two each alternating C-myc NLS and a 12aa linker at the 3' end. In some embodiments, an optimized nNme2Cas9-ABEmax construct further comprises at least two each alternating 15aa linker and C-myc NLS at the 5' end. See, Figure 1E for example.

In one embodiment, an optimized nNme2Cas9-ABEmax construct has the following amino acid sequence:

PAAKRVKLD***GGSGGGSGGGSG***PAAKRVKLD***GGSGGGSGGGSGPLE*****PKKKRKV**
<u>SEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHA</u>
<u>EIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGS</u>
<u>LMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD</u>***<u>SGGSSGGS</u>***
<u>SGSETPGTSESATPESSGGSSGGS</u>**<u>SEVEFSHEYWMRHALTLAKRARDEREVPVGAVL</u>**
<u>VLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVM</u>

CAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAAL LCYFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGSMAA FKPNPINYDIDKL*AAFKPNPINYILGLAIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPK TGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAA ALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPA ELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMT QRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERAT LMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLK DKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRI VPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRY GSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLY EQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYE YFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVA DHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRF VRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEK LRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEI KLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKT QESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRI DDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLV LIQKYQVNELGKEIRPCRLKKRPPVRVYPYDVPDYAGYPYDVPDYAGSYPYDVPDYAGSAA PAAKKKKLDFESGEFLQPGGSTSSRGSAA**PAAKRVKLD***GGSGGGSGGGSGSG***PAAKRV KLD**

hTadA7.10 (underlined), hTadA*7.10 (underlined/bold), linker *(bold italics),* nNme2Cas9 (*italics*), Cmyc-NLS (plain), SV40-NLS (bold).

In some embodiments, a plasmid nSpCas9-ABEmax (Addgene ID:112095) was used for experimental controls and for molecular cloning. In some embodiments, a plasmid nSpCas9-CBE4 (Addgene ID: 100802) was used for experimental controls and for molecular cloning.

Electroporation of HEK293T cells with DNA plasmids comprising a YE1-BE3-Nme2Cas9 nucleotide deaminase fusion protein achieved robust single-base editing of a C•G base pair to a T•A base pair at an endogenous target site (TS25). See, Figures 2A-C.

Figure 2 presents exemplary data of the electroporation of HEK293T cells with DNA plasmids comprising a YE1-BE3-nNme2Cas9 (D16A)-UGI fusion protein efficiently converting C to T at endogenous target site 25 (TS25) in HEK293T cells via nucleofection. Figure 2A shows exemplary sequences for a TS25 endogenous target site (within the black rectangle). GN23 sgRNA base-pairs with the target DNA strand, leaving the displaced DNA strand for cytidine deaminase to edit (e.g. new green nucleotides). Figure 2B shows exemplary sequencing data showing a doublet nucleotide peak (7th position from 5' end; arrow) demonstrating the successful single base editing of a cytidine to a thymidine (e.g., a C•G base pair conversion to a T•A base pair). Figure 2C shows an exemplary quantitation of the data shown in Figure 2B plotting the percent conversion of C → T single base editing. The percentage of C converted to T is about 40% in the base editor- and sgRNA-treated sample (p-value = 6.88 x 10-6). The "no sgRNA" control displays the background noise due to Sanger sequencing. EditR (Kluesner et al., 2018) was used to perform the analysis.

Four other YE1-BE3-nNme2Cas9/D16A mutant fusion proteins were co-expressed with enhanced green fluorescent protein (EGFP) in a stable K562-derived cell line expressing enhanced green fluorescent protein (EGFP). Each YE1-BE3-nNme2Cas9/D16A mutant fusion protein had a specific UGI target site. See, Figures 3A-D.

Deep-sequencing analysis indicates YE1-BE3-nNme2Cas9 converts C residues to T residues at each of the four EGFP target sites. The percentage of editing ranged from 0.24% to 2%. The potential base editing window is from nucleotides 2-8 in the displaced DNA strand, counting the nucleotide at the 5' (PAM-distal) end as nucleotide #1. See, Figures 3A-D.

Figure 3 presents exemplary specific UGI target sites that were respectively integrated into YE1-BE3-nNme2Cas9/D16A mutant fusion proteins and co-expressed with enhanced green fluorescent protein (EGFP) in a stable K562-derived cell line. Converted bases are highlighted in orange color. Background signals were filtered using negative control samples (YE1-BE3-nNme2Cas9 nucleofected K562 cells without sgRNA constructs). N_4CC PAMs are boxed. The percentage of total reads exhibiting mutations in base-editor-targeted sites is shown in the right

column. Figure 3A shows an exemplary EGFP-Site 1. Figure 3B shows an exemplary EGFP-Site 2. Figure 3C shows an exemplary EGFP-Site 3. Figure 3D shows an exemplary EGFP-Site 4.

Electroporation of HEK293T cells with DNA plasmids comprising a YE1-BE3-nNme2Cas9 *c-fos* promoter achieved robust single-base editing of a C•G base pair to a T•A base pair at endogenous target sites in the *c-fos* promoter (Figure 3E). Figure 3E shows an exemplary deep-sequencing analysis indicating where YE1-BE3-nNme2Cas9 converts C residues to T residues at endogenous *c-fos* promoter region. The percentage of total reads exhibiting mutations in base-editor-targeted sites is shown in the right column. The converted bases are highlighted in orange or yellow color. Background signals were filtered using negative control samples. The highest percentage of editing is 32.50%. Figure 3F shows an exemplary deep-sequencing analysis indicating where ABE7.10-nNme2Cas9 or ABEmax (Koblan *et al.,* 2018)-nNme2Cas9 converts A residues to G residues at endogenous *c-fos* promoter region. The percentage of total reads exhibiting mutations in base-editor-targeted sites is shown in the right column. The converted bases are highlighted in orange color. Background signals were filtered using negative control samples. The percentage of editing is 0.53% by ABE7.10-nNme2Cas9 or 2.33% by ABEmax-nNme2Cas9 (D16A).

In one embodiment, the present invention contemplates the expression of an ABE7.10-nNme2Cas9 (D16A) fusion protein for base editing. Although it is not necessary to understand the mechanism of an invention, it is believed that Nme2Cas9 base editing may be an effective treatment for tyrosinemia by reversing a G-to-A point mutation in the *Fah* gene with an ABE7.10-nNme2Cas9 (D16A) fusion protein.

G-to-A mutation (red) at the last nucleotide of exon 8 in *Fah* gene, causing exon skipping. FAH deficiency leads to toxin accumulation and severe liver damage. The position of a SpyCas9 PAM (black rectangular box) downstream of the mutation is not optimal for designing the sgRNA since the A mutation is out of the efficient base editing window of ABE7.10, which is 4-7th nt at the 5' (PAM-distal) end (underlined) (Gaudelli et al., 2017).

However, there are two Nme2Cas9 PAMs (red rectangular box) in the downstream sequences that can potentially correct the mutation and revert DNA sequence to wildtype via ABE7.10-nNme2Cas9 (D16A). See, Figure 4.

Figure 4 presents an exemplary alignment of the wildtype *Fah* gene with the tyrosinemia *Fah* mutant gene showing an A-G single base gene editing target site (position 9). The respective

SpyCas9 single PAM site and NmeCas9 double PAM sites are indicated for demonstrating the suboptimal targeting window relative to the SpyCas9 PAM site. This figure serves as a potential example of a site where Nme2Cas9 could overcome limitations of existing base editors. It is further believed that the NmeCas9 base editor described herein can perform precise base editing that cannot be achieved with conventional SpyCas9-derived base editors due to a suboptimal base editing window relative to available PAMs nearby.

Furthermore, we contemplate extending base editing to a tyrosinemia mouse model for reversing the G-to-A point mutation by viral delivery methods using ABEmax-nNme2Cas9 (D16A), where the desired editing cannot be achieved with SpyCas9-derived base editors due to a suboptimal base editing window relative to available PAMs nearby (e.g. Figure 4).

B. NmeCas9 Constructs: Compact & Hyperaccurate

Clustered, regularly interspaced, short, palindromic repeats (CRISPR) along with CRISPR-associated (Cas) proteins constitute bacterial and archaeal adaptive immune pathways against phages and other mobile genetic elements (MGEs) (Barrangou et al., 2007; Brouns et al., 2008; Marraffini and Sontheimer, 2008). In Type II CRISPR systems, CRISPR RNA (crRNA) is bound to a trans-activating crRNA (tracrRNA) and loaded onto a Cas9 effector protein that cleaves MGE nucleic acids complementary to the crRNA (Garneau et al., 2010; Deltcheva et al., 2011; Sapranauskas et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012). The crRNA:tracrRNA hybrid can be fused into a single-guide RNA (sgRNA) (Jinek et al., 2012). The RNA programmability of Cas9 endonucleases has made it a powerful genome editing platform in biotechnology and medicine (Cho et al., 2013; Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2013; Mali et al., 2013b).

In addition to sgRNA, Cas9 target recognition is usually associated with a 1-5 nucleotide signature downstream of the complementary DNA sequence, called a protospacer adjacent motif (PAM) (Deveau et al., 2008; Mojica et al., 2009). Cas9 orthologs exhibit considerable diversity in PAM length and sequence. Among Cas9 orthologs that have been characterized, *Streptococcus pyogenes* Cas9 (SpyCas9) is the most widely used, in part because it recognizes a short NGG PAM (Jinek et al., 2012) (N represents any nucleotide) that affords a high density of targetable sites. Nevertheless, Spy's relatively large size (i.e., 1,368 amino acids) makes this Cas9 difficult to package (along with sgRNA and promoters) into a single recombinant adeno-associated virus (rAAV). This has been shown to be a drawback for therapeutic applications

given the promise shown by AAV vectors for in vivo gene delivery (Keeler et al., 2017). Moreover, SpyCas9 and its RNA guides have required extensive characterization and engineering to minimize the tendency to edit near-cognate, off-target sites. (Bolukbasi et al., 2015b; Tsai and Joung, 2016; Tycko et al., 2016; Chen et al., 2017; Casini et al., 2018; Yin et al., 2018). To date, subsequent engineering efforts have not overcome these size limitations.

Several Cas9 orthologs of less than 1,100 amino acids in length obtained from diverse species have been validated for mammalian genome editing, including strains of *N. meningitidis* (NmeCas9, 1,082 aa) (Esvelt et al., 2013; Hou et al., 2013), *Staphylococcus aureus* (SauCas9, 1,053 aa) (Ran et al., 2015), *Campylobacter jejuni* (CjeCas9, 984 aa) (Kim et al., 2017), and *Geobacillus stearothermophilus* (GeoCas9, 1,089 aa) (Harrington et al., 2017b). NmeCas9, CjeCas9, and GeoCas9 are representatives of type II-C Cas9s (Mir et al., 2018), most of which are <1,100 aa. With the exception of GeoCas9, each of these shorter sequence orthologs has been successfully deployed for in vivo editing via all-in-one AAV delivery (in which a single vector expresses both guide and effector) (Ran et al., 2015; Kim et al., 2017; Ibraheim et al., 2018, submitted). Furthermore, NmeCas9 and CjeCas9 have been shown to be naturally resistant to off-target editing (Lee et al., 2016; Kim et al., 2017; Amrani et al., 2018, submitted). However, the PAMs that are recognized by compact Cas9s are usually longer than that of SpyCas9, substantially reducing the number of targetable sites at or near a given locus; for example, i) N_4GAYW/N_4GYTT/N_4GTCT for NmeCas9 (Esvelt et al., 2013; Hou et al., 2013; Lee et al., 2016; Amrani et al., 2018); ii) N_2GRRT for SauCas9 (Ran et al., 2015); iii) N_4RYAC for CjeCas9 (Kim et al., 2017); and iv) N_4CRAA/N_4GMAA for GeoCas9s (Harrington et al., 2017b) (Y = C, T; R = A, G; M = A, C; W = A, T). A smaller subset of target sites is advantageous for highly accurate and precise gene editing tasks including, but not limited to: i) editing of small targets (e.g. miRNAs); ii) correction of mutations by base editing which alters a very narrow window of bases relative to the PAM (Komor et al., 2016; Gaudelli et al., 2017); or iii) precise editing via homology-directed repair (HDR) which is most efficient when the rewritten bases are close to the cleavage site (Gallagher and Haber, 2018). Because of PAM restrictions, many editing sites cannot be targeted using all-in-one AAV vectors for in vivo delivery even with these shorter Cas9 proteins. For example, A SauCas9 mutant (SauCas9^{KKH}) has been developed that has reduced PAM constraints (N_3RRT), though this increase in targeting

range often comes at the cost of reduced on-target editing efficacy, and off-target edits are still observed. (Kleinstiver et al., 2015).

Safe and effective CRISPR-based therapeutic gene editing will be greatly enhanced by Cas9 orthologs and variants that are highly active in human cells, resistant to off-targeting, sufficiently compact for all-in-one AAV delivery, and capable of accessing a high density of genomic sites. In one embodiment, the present invention contemplates a compact, hyper-accurate Cas9 (Nme2Cas9) from a distinct strain of *N. meningitidis*. In one embodiment, the present invention contemplates a method for single-AAV delivery of Nme2Cas9 and its sgRNA to perform efficient genome editing in vivo and/or ex vivo. Although it is not necessary to understand the mechanism of an invention, it is believed that this ortholog functions efficiently in mammalian cells and recognizes an N_4CC PAM that affords a target site density identical to that of wild-type SpyCas9 (e.g., every 8 bp on average, when both DNA strands are considered).

1. PAM Interacting Domains And Anti-CRISPR Proteins

PAM recognition by Cas9 orthologs occurs predominantly through protein-DNA interactions between the PAM Interacting Domain (PID) and the nucleotides adjacent to the protospacer (Jiang and Doudna, 2017). PAM mutations often enable phage escape from type II CRISPR immunity (Paez-Espino et al., 2015), placing these systems under selective pressure not only to acquire new CRISPR spacers, but also to evolve new PAM specificities via PID mutations. In addition, some phages and MGEs express anti-CRISPR (Acr) proteins that inhibit Cas9 (Pawluk et al., 2016; Hynes et al., 2017; Rauch et al., 2017). PID binding is an effective inhibitory mechanism adopted by some Acrs (Dong et al., 2017; Shin et al., 2017; Yang and Patel, 2017), suggesting that PID variation may also be driven by selective pressure to escape Acr inhibition. Cas9 PIDs can evolve such that closely-related orthologs recognize distinct PAMs, as illustrated recently in two species of *Geobacillus*. The Cas9 encoded by *G. stearothermophilus* recognizes a N_4CRAA PAM, but when its PID was swapped with that of strain LC300's Cas9, its PAM requirement changed to N_4GMAA (Harrington et al., 2017b).

In one embodiment, the present invention contemplates a plurality of *N. meninigitidis* Cas9 orthologs with divergent PIDs that recognize different PAMs. In one embodiment, the present invention contemplates a Cas9 protein with a high sequence identity (>80% along their entire lengths) to that of NmeCas9 strain 8013 (Nme1Cas9) (Zhang et al., 2013). Nme1Cas9 also has a small size and naturally high accuracy as discussed above. (Lee et al., 2016; Amrani et al.,

2018). Alignments revealed three clades of meningococcal Cas9 orthologs, each with >98% identity in the N-terminal ~820 amino acid (aa) residues, which includes all regions of the protein other than the PID. See, Figure 5A and Figure 6A.

All of these Cas9 orthologs are 1,078-1,082 aa in length. The first clade (group 1) includes orthologs in which the >98% aa sequence identity with Nme1Cas9 extends through the PID. In contrast, the other two groups had PIDs that were significantly diverged from that of Nme1Cas9, with group 2 and group 3 orthologs averaging ~52% and ~86% PID sequence identity with Nme1Cas9, respectively. One meningococcal strain was selected from each group: i) De11444 from group 2; and ii) 98002 from group 3 for detailed analysis, which are referred to herein as Nme2Cas9 (1,082 aa) and Nme3Cas9 (1,081 aa), respectively. The CRISPR-*cas* loci from these two strains have repeat sequences and spacer lengths that are identical to those of strain 8013. See, Figure 6B. This strongly suggested that their mature crRNAs also have 24nt guide sequences and 24 nt repeat sequences (Zhang et al., 2013). Similarly, the tracrRNA sequences of De11444 and 98002 were 100% identical to the 8013 tracrRNA. See, Figure 6B. These observations imply that the same sgRNA sequence scaffold can guide DNA cleavage by all three Cas9s.

To determine whether these Cas9 orthologs have distinct PAMs, the PID of Nme1Cas9 was replaced with that of either Nme2Cas9 or Nme3Cas9. To identify the corresponding PAM requirements, these protein chimeras were expressed in *Escherichia coli*, purified, and used for in vitro PAM identification (Karvelis et al., 2015; Ran et al., 2015; Kim et al., 2017). Briefly, a pool of DNA fragments containing a protospacer followed by a 10-nt randomized sequence was cleaved in vitro using recombinant Cas9 and a cognate, in vitro-transcribed sgRNA. See, Figure 5B. Only those DNAs containing a Cas9 PAM sequence were expected to be cleaved. Cleavage products were then sequenced to identify the PAMs. See, Figures 5C-D.

The expected N_4GATT PAM consensus was validated in the recovered full-length Nme1Cas9. See, Figure 5C. Chimeric PID-swapped derivatives exhibited a strong preference for a C residue in the 5th position in place of the G recognized by Nme1Cas9. See, Figure 5D.

In one embodiment, ABE7.10-nNme2Cas9 (D16A) is used for single-base editing of A•T base pair to a G•C base pair. In one embodiment, BEmax-nNme2Cas9 (D16A) is used for single-base editing of A•T base pair to a G•C base pair. (See, Figure 3F).

Figure 5 illustrates exemplary three closely related *Neisseria meningitidis* Cas9 orthologs that have distinct PAMs. Figure 5A shows an exemplary schematic showing mutated residues (orange spheres) between Nme2Cas9 (left) and Nme3Cas9 (right) mapped onto the predicted structure of Nme1Cas9, revealing the cluster of mutations in the PID (black). Figure 5B shows an exemplary experimental workflow of the in vitro PAM discovery assay with a 10-bp randomized PAM region. Following in vitro digestion, adapters were ligated to cleaved products for library construction and sequencing. Figure 5C shows exemplary sequence logos resulting from in vitro PAM discovery reveal the enrichment of a N_4GATT PAM for Nme1Cas9, consistent with its previously established specificity. Figure 5D shows exemplary sequence logos indicating that Nme1Cas9 with its PID swapped with that of Nme2Cas9 (left) or Nme3Cas9 (right) requires a C at PAM position 5. The remaining nucleotides were not determined with high confidence due to the modest cleavage efficiency of the PID-swapped protein chimeras (see Figure 6C). Figure 5E shows an exemplary sequence logo showing that full-length Nme2Cas9 recognizes an N_4CC PAM, based on efficient substrate cleavage of a target pool with a fixed C at PAM position 5, and with PAM nts 1-4 and 6-8 randomized.

Any remaining PAM nucleotides could not be confidently assigned due to the low cleavage efficiencies of the chimeric proteins under the conditions used. See, Figure 6C. To further resolve the PAMs, in vitro assays were performed on a library with a 7-nt randomized sequence possessing an invariant C at the 5th PAM position (e.g., 5'-NNNNCNNN-3' on the sgRNA-noncomplementary strand). This strategy yielded a much higher cleavage efficiency and the results indicated that the Nme2Cas9 and Nme3Cas9 PIDs recognize NNNNCC(A) and NNNNCAAA PAMs, respectively. See, Figures 6C-D. The Nme3Cas9 consensus is similar to that of GeoCas9 (Harrington et al., 2017b).

These tests were repeated using a full-length Nme2Cas9 (rather than a PID-swapped chimera) with the NNNNCNNN DNA pool, and again a NNNNCC(A) consensus was recovered. See, Figure 5E. It was noted that this test had more efficient cleavage. See, Figure 6C. These data suggest that one or more of the 15 amino acid changes in Nme2Cas9 (relative to Nme1Cas9) outside of the PID support efficient DNA cleavage activity. See, Figure 6C. Because the unique, 2-3 nt PAM of Nme2Cas9 affords a higher density of potential target sites than the previously described compact Cas9 orthologs, it was selected for further analyses.

Figure 6 presents a characterization of *Neisseria meningitidis* Cas9 orthologs with rapidly-evolving PIDs, as related to Figure 5**.** Figure 6A shows an exemplary unrooted phylogenetic tree of NmeCas9 orthologs that are >80% identical to Nme1Cas9. Three distinct branches emerged, with the majority of mutations clustered in the PID. Groups 1 (blue), 2 (orange), and 3 (green) have PIDs with >98%, approximately 52%, and approximately86% identity to Nme1Cas9, respectively. Three representative Cas9 orthologs (one from each group) (Nme1Cas9, Nme2Cas9 and Nme3Cas9) are indicated. Figure 6B shows an exemplary schematic showing the CRISPR-*cas* loci of the strains encoding the three Cas9 orthologs (Nme1Cas9, Nme2Cas9, and Nme3Cas9) from (A). Percent identities of each CRISPR-Cas component with *N. meningitidis* 8013 (encoding Nme1Cas9) are shown. Blue and red arrows denote pre-crRNA and tracrRNA transcription initiation sites, respectively. Figure 6C shows an exemplary normalized read counts (% of total reads) from cleaved DNAs from the in vitro assays for intact Nme1Cas9 (grey), for chimeras with Nme1Cas9's PID swapped with those of Nme2Cas9 and Nme3Cas9 (mixed colors), and for full-length Nme2Cas9 (orange), are plotted. The reduced normalized read counts indicate lower cleavage efficiencies in the chimeras. Figure 6D shows an exemplary sequence logos from the in vitro PAM discovery assay on an NNNNCNNN PAM pool by Nme1Cas9 with its PID swapped with those of Nme2Cas9 (left) or Nme3Cas9 (right).

2. N_4CC PAM-Directed Gene Editing

To test the efficacy of Nme2Cas9 in human genome editing, a full-length (e.g., not PID-swapped) human-codon-optimized Nme2Cas9 construct was cloned into a mammalian expression plasmid with appended nuclear localization signals (NLSs) and linkers validated previously for Nme1Cas9 (Amrani et al., 2018). For initial tests, a modified, fluorescence-based Traffic Light Reporter (TLR2.0) was used (Certo et al., 2011). Briefly, a disrupted GFP is followed by an out-of-frame T2A peptide and mCherry cassette. When DNA double-strand breaks (DSBs) are introduced in the broken-GFP cassette, a subset of non-homologous end joining (NHEJ) repair events leave +1-frameshifted indels, placing mCherry in frame and yielding red fluorescence that can be easily quantified by flow cytometry See, Figure 7A. Homology-directed repair (HDR) outcomes can also be scored simultaneously by including a DNA donor that restores the functional GFP sequence, yielding a green fluorescence (Certo et al., 2011). Because some indels do not introduce a +1 frameshift, the fluorescence readout

generally provides an underestimate of the true editing efficiency. Nonetheless, the speed, simplicity, and low cost of the assay makes it useful as an initial, semi-quantitative measure of genome editing in HEK293T cells carrying a single TLR2.0 locus incorporated via lentivector.

For initial tests, Nme2Cas9 plasmid was transiently co-transfected with one of fifteen sgRNA plasmids carrying spacers that target TLR2.0 sites with N_4CC PAMs. No HDR donor was included, so only NHEJ-based editing (mCherry) was scored. Most sgRNAs were in a G23 format (i.e. a 5'-terminal G to facilitate transcription, followed by a 23nt guide sequence), as used routinely for Nme1Cas9 (Lee et al., 2016; Pawluk et al., 2016; Amrani et al., 2018; Ibraheim et al., 2018). No sgRNA and an sgRNA targeting an N_4GATT PAM were used as negative controls, and SpyCas9+sgRNA and Nme1Cas9+sgRNA co-transfections (targeting NGG and N_4GATT protospacers, respectively) were included as positive controls. Editing by SpyCas9 and Nme1Cas9 was readily detectable (~28% and 10% mCherry, respectively). See, Figure 7B.

For Nme2Cas9, all 15 targets with N_4CC PAMs were functional, though to various extents ranging from 4% to 20% mCherry. These fifteen sites include examples with each of the four possible nucleotides in the 7th PAM position (e.g., after the CC dinucleotide), indicating that a slight preference for an A residue was observed in vitro (Figure 5E) does not reflect a PAM requirement for editing applications in human cells. The N_4GATT PAM control yielded mCherry signal similar to no-sgRNA control. See, Figure 7B.

To determine whether both C residues in the N_4CC PAM are involved in editing, a series of N_4DC (D = A, T, G) and N_4CD PAM sites were tested in TLR2.0 reporter cells. See, Figures 8A and 8B. No detectable editing was found at any of these sites, providing an initial indication that both C residues of the N_4CC PAM consensus are required for efficient Nme2Cas9 activity.

The length of the spacer in the crRNA differs among Cas9 orthologs and can affect on- vs. off-target activity (Cho et al., 2014; Fu et al., 2014). SpyCas9's optimal spacer length is 20 nts, with truncations down to 17 nts tolerated (Fu et al., 2014). In contrast, Nme1Cas9 usually has 24-nt spacers (Hou et al., 2013; Zhang et al., 2013), and tolerates truncations down to 18-20 nts (Lee et al., 2016; Amrani et al., 2018). To test spacer length requirements for Nme2Cas9, guide RNA plasmids were created for each targeted single TLR2.0 site, but with varying spacer lengths. See, Figure 7C and Figure 8C. Comparable activities were observed with G23, G22 and G21 guides, but significantly decreased activity upon further truncation to G20 and G19 lengths.

See, Figure 7C. These results validate Nme2Cas9 as a genome editing platform, with 22-24 nt guide sequences, at N_4CC PAM sites in cultured human cells.

Figure 7 presents exemplary data showing that Nme2Cas9 uses a 22-24 nt spacer to edit sites adjacent to an N_4CC PAM. All experiments were done in triplicate, and error bars represent the standard error of the mean (s.e.m.). Figure 7A shows an exemplary schematic diagram depicting transient transfection and editing of HEK293T TLR2.0 cells, with mCherry+ cells detected by flow cytometry 72 hours after transfection. Figure 7B shows an exemplary Nme2Cas9 editing of the TLR2.0 reporter. Sites with N_4CC PAMs were targeted with varying efficiencies, while no Nme2Cas9 targeting was observed at an N_4GATT PAM or in the absence of sgRNA. SpyCas9 (targeting a previously validated site with an NGG PAM) and Nme1Cas9 (targeting N_4GATT) were used as positive controls. Figure 7C shows an exemplary effect of spacer length on the efficiency of Nme2Cas9 editing. An sgRNA targeting a single TLR2.0 site, with spacer lengths varying from 24 to 20 nts (including the 5'-terminal G required by the U6 promoter), indicate that highest editing efficiencies are obtained with 22-24 nt spacers. Figure 7D shows an exemplary An Nme2Cas9 dual nickase can be used in tandem to generate NHEJ- and HDR-based edits in TLR2.0. Nme2Cas9- and sgRNA-expressing plasmids, along with an 800-bp dsDNA donor for homologous repair, were electroporated into HEK293T TLR2.0 cells, and both NHEJ (mCherry+) and HDR (GFP+) outcomes were scored by flow cytometry. HNH nickase, Nme2Cas9^{D16A}; RuvC nickase, Nme2Cas9^{H588A}. Cleavage sites 32 bp and 64 bp apart were targeted using either nickase. The HNH nickase (Nme2Cas9^{D16A}) yielded efficient editing, particularly with the cleavage sites that were separated by 32 bp, whereas the RuvC nickase (Nme2Cas9^{H588A}) was not effective. Wildtype Nme2Cas9 was used as a control.

3. Precise Editing By HDR And HNH Nickase

Cas9 enzymes use their HNH and RuvC domains to cleave the guide-complementary and non-complementary strand of the target DNA, respectively. SpyCas9 nickases (nCas9s), in which either the HNH or RuvC domain is mutationally inactivated, have been used to induce homology-directed repair (HDR) and to improve genome editing specificity via DSB induction by dual nickases (Mali et al., 2013a; Ran et al., 2013).

To test the efficacy of Nme2Cas9 as a nickase, a Nme2Cas9^{D16A} (HNH nickase) and Nme2Cas9^{H588A} (RuvC nickase) were created, which possess alanine mutations in catalytic residues of the RuvC and HNH domains, respectively (Esvelt et al., 2013; Hou et al., 2013;

Zhang et al., 2013). TLR2.0 cells, along with a GFP donor dsDNA, were used to determine whether Nme2Cas9-induced nicks can induce precise edits via HDR. Target sites within TLR2.0 were used to test the functionality of each nickase using guides targeting cleavage sites spaced 32 bp and 64 bp apart. See, Figure 7D. Wildtype Nme2Cas9 targeting a single site showed efficient editing, with both NHEJ and HDR as outcomes of repair. For nickases, cleavage sites 32 bp and 64 bp apart showed editing using the Nme2Cas9^{D16A} (HNH nickase), but neither target pair worked with Nme2Cas9^{H588A}. These results suggest that Nme2Cas9 HNH nickase can be used for efficient genome editing, as long as the sites are in close proximity.

Studies in previously characterized Cas9s have identified a specific region proximal to the PAM where Cas9 activity is highly sensitive to sequence mismatches. This 8 to 12-nt region is known as the seed sequence and has been observed among all Cas9s characterized to date (Gorski et al., 2017). To determine whether Nme2Cas9 also possesses a seed sequence, a series of transient transfections was performed, each targeting the same locus in TLR2.0, but with a single-nucleotide mismatch at different positions of the guide. See, Figure 8D. A significant decrease in the number of mCherry-positive cells was observed for mismatches in the first 10-12 nts proximal to the PAM, suggesting that Nme2Cas9 possesses a seed sequence in this region.

Figure 8 presents exemplary data showing PAM, spacer, and seed requirements for Nme2Cas9 targeting in mammalian cells, as related to Figure 7. All experiments were done in triplicate and error bars represent s.e.m. Figure 8A shows an exemplary Nme2Cas9 targeting at $N_4C\underline{D}$ sites in TLR2.0, with editing estimated based on mCherry+ cells. Four sites for each non-C nucleotide at the tested position ($N_4C\underline{A}$, $N_4C\underline{T}$ and $N_4C\underline{G}$) were examined, and an N_4CC site was used as a positive control. Figure 8B shows an exemplary Nme2Cas9 targeting at $N_4\underline{D}C$ sites in TLR2.0 [similar to (A)]. Figure 8C shows exemplary guide truncations on a TLR2.0 site (distinct from that in Figure 2C) with a N_4CCA PAM, revealing similar length requirements as those observed at the other site. Figure 8D shows exemplary Nme2Cas9 targeting efficiency is differentially sensitive to single-nucleotide mismatches in the seed region of the sgRNA. Data show the effects of walking single-nucleotide sgRNA mismatches along the 23-nt spacer in a TLR2.0 target site.

4. Delivery Methods To Mammalian Cell Types

Nme2Cas9's ability to function in different mammalian cell lines was tested using various delivery methods. As an initial test, forty (40) different sites (29 with a N_4CC PAM, and

11 sites were tested with a N_4CD PAM). Several loci were selected (*AAVS1*, *VEGFA*, etc.), and target sites with N_4CC PAMs were randomly chosen for editing with Nme2Cas9. Editing (%) was determined by transiently transfecting 150 ng of Nme2Cas9 along with 150 ng of sgRNA plasmids followed by TIDE analysis 72 hours post-transfection. A subset of sites exhibiting a range of editing efficiencies in this initial screen was selected for repeat analyses in triplicate. See, Figure 9A; and Table 1.

Figure 9 presents exemplary data showing Nme2Cas9 genome editing at endogenous loci in mammalian cells via multiple delivery methods. All results represent 3 independent biological replicates, and error bars represent s.e.m. Figure 9A shows an exemplary Nme2Cas9 genome editing of endogenous human sites in HEK293T cells following transient transfection of Nme2Cas9- and sgRNA-expressing plasmids. 40 sites were screened initially (Table 1); the 14 sites shown (selected to include representatives of varying editing efficiencies, as measured by TIDE) were then re-analyzed in triplicate. An Nme1Cas9 target site (with an N_4GATT PAM) was used as a negative control. Figure 9B shows exemplary data charts: Left panel: Transient transfection of a single plasmid expressing both Nme2Cas9 and sgRNA (targeting the *Pcsk9* and *Rosa26* loci) enables editing in Hepa1-6 mouse cells, as detected by TIDE. Right panel: Electroporation of sgRNA plasmids into K562 cells stably expressing Nme2Cas9 from a lentivector results in efficient indel formation. Figure 9C shows exemplary Nme2Cas9 can be electroporated as an RNP complex to induce genome editing. 40 picomoles Cas9 along with 50 picomoles of in vitro-transcribed sgRNAs targeting three different loci were electroporated into HEK293T cells. Indels were measured after 72h using TIDE.

Table 1. Exemplary Endogenous human genome editing sites targeted by Nme2Cas9.

Number	Site Name	Spacer Seq	PAM	Locus	Editing (%)	TIDE Primer name	FW TIDE primer	RV TIDE primer
1	TS1	GGTTCTGGGTACTTTTATCTGTCC	CCTCCACC	*AAVS1*	ND	AAVS1_TIDE1	TGGCTTAGCACCTCTCCAT	AGAACTCAGGACCAACTTATTCTG
2	TS4	GTCTGCCTAACAGGAGGTGGGGGT	TAGACGAA	*AAVS1*	11	AAVS1_TIDE1	TGGCTTAGCACCTCTCCAT	AGAACTCAGGACCAACTTATTCTG
3	TS5	GAATATCAGGAGACTAGGAAGGAG	GAGGCCTA	*AAVS1*	15	AAVS1_TIDE1	TGGCTTAGCACCTCTCCAT	AGAACTCAGGACCAACTTATTCTG
4	TS6	GCCTCCCTGCAGGGCTGCTCCC	CAGCCCAA	*LINC01588*	20	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
5	TS10	GAGCTAGTCTTCTTCCTCCAACCC	GGGCCCTA	*AAVS1*	3.5	AAVS1_TIDE1	TGGCTTAGCACCTCTCCAT	AGAACTCAGGACCAACTTATTCTG
6	TS11	GATCTGTCCCCTCCACCCCACAGT	GGGGCCAC	*AAVS1*	9	AAVS1_TIDE1	TGGCTTAGCACCTCTCCAT	AGAACTCAGGACCAACTTATTCTG
7	TS12	GGCCCAAATGAAAGGAGTGAGAGG	TGACCCGA	*AAVS1*	10	AAVS1_TIDE2	TCCGTCTTCCTCCACTCC	TAGGAAGGAGGAGGCCTAAG
8	TS13	GCATCCTCTTGCTTTCTTTGCCTG	GACACCCC	*AAVS1*	2	AAVS1_TIDE2	TCCGTCTTCCTCCACTCC	TAGGAAGGAGGAGGCCTAAG
9	TS16	GGAGTCGCCAGAGGCCGGTGGTGG	ATTTCCTC	*LINC01588*	28	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
10	TS17	GCCCAGCGGCCGGATATCAGCTGC	CACGCCCG	*LINC01588*	ND	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
11	TS18	GGAAGGGAACATATTACTATTGC	TTTCCCTC	*CYBB*	1	NTS55_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
12	TS19	GTGGAGTGGCCTGCTATCAGCTAC	CTATCCAA	*CYBB*	6	NTS55_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
13	TS20	GAGGAAGGGAACATATTACTATTG	CTTTCCCT	*CYBB*	11.2	NTS55_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
14	TS21	GTGAATTCTCATCAGCTAAAATGC	CAAGCCTT	*CYBB*	1	NTS55_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
15	TS25	GCTCACTCACCCACACAGACACAC	ACGTCCTC	*VEGF A*	15.6	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
16	TS26	GGAAGAATTTCATTCTGTTCTCAG	TTTTCCTG	*CFTR*	2	hCFTR_TIDE1	GGTGATTATGGGAGAACTGGAG	ACCATTGAGGACGTTTGTCTCAC
17	TS27	GCTCAGTTTTCCTGGATTATGCCT	GGCACCAT	*CFTR*	4	hCFTR_TIDE1	GGTGATTATGGGAGAACTGGAG	ACCATTGAGGACGTTTGTCTCAC
18	TS31	GCGTTGGAGCGGGGAGAAGGCCAG	GGGTCACT	*VEGFA*	9	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
19	TS34	GGGCCGCGGAGATAGCTGCAGGGC	GGGGCCCC	*LINC01588*	ND	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
20	TS35	GCCCACCCGGCGGCGCCTCCCTGC	AGGGCTGC	*LINC01588*	ND	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
21	TS36	GCGTGGCAGCTGATATCCGGCCGC	TGGGCGTC	*LINC01588*	ND	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
22	TS37	GCCGCGGGCGCGACGTGGAGCCAGC	CCCGCAAA	*LINC01588*	ND	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
23	TS38	GTGCTCCCCAGCCCAAACCGCCGC	GGCGCGAC	*LINC01588*	2	LINC01588_TIDE	AGGAGCCTTCTGACTGCTGCA	ATGACAGACACAACCAGAGGGCA
24	TS41	GTCAGATTGGCTTGCTCGGAATTG	CCAGCCAA	*AGA*	3	AGA_TIDE1	GGCATAAGGAAATCGAAGGTC	CATGTCCTCAAGTCAAGAACAAG
25	TS44	GCTGGGTGAATGGAGCGAGCAGCG	TCTTCGAG	*VEGFA*	3	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
26	TS45	GTCCTGGAGTGACCCCTGGCCTTC	TCCCCGCT	*VEGFA*	7.4	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
27	TS46	GATCCTGGAGTGACCCCTGGCCTT	CTCCCCGC	*VEGFA*	6	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
28	TS47	GTGTGTCCCTCTCCCCACCCGTCC	CTGTCCGG	*VEGFA*	23.1	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
29	TS48	GTTGGAGCGGGGAGAAGGCCAGGG	GTCACTCC	*VEGFA*	2	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
30	TS49	GCGTTGGAGCGGGGAGAAGGCCAG	GGGTCACT	*VEGFA*	4	VEGF_TIDE3	TACATGAAGCAACTCCAGTCCC	ATCAAATTCCAGCACCGAGCGC
31	TS50	GTACCCTCCAATAATTTGGCTGGC	AATTCCGA	*AGA*	6	AGA_TIDE1	GGCATAAGGAAATCGAAGGTC	CATGTCCTCAAGTCAAGAACAAG
32	TS51	GATAATTTGGCTGGCAATTCCGAG	CAAGCCAA	*AGA*	4.5	AGA_TIDE1	GGCATAAGGAAATCGAAGGTC	CATGTCCTCAAGTCAAGAACAAG
33	TS58 (DS1)	GCAGGGGCCAGGTGTCCTTCTCTG	GGGGCCTC	*VEGFA*	5	VEGF_TIDE4	CACGGGCAGCATGGGAATAGT	GCTAGGGGAGAGTCCCACTGTCCA
34	TS59 (DS2)	GAATGGCAGGCGGAGGTTGTACTG	GGGGCCAG	*VEGFA*	11.5	VEGF_TIDE5	CTGTGTGGCTTTGCTTTGGTC	GTAGGGTGTGATGGGAGGCTAAGC
35	TS60 (DS3)	GAGTGAGAGAGTGAGAGAGAGACA	CGGGCCAG	*VEGFA*	3	VEGF_TIDE5	CTGTGTGGCTTTGCTTTGGTC	GTAGGGTGTGATGGGAGGCTAAGC
36	TS61 (DS4)	GTGAGCAGGCACCTGTGCCAACAT	GGGCCGCC	*VEGFA*	3.5	VEGF_TIDE5	CTGTGTGGCTTTGCTTTGGTC	GTAGGGTGTGATGGGAGGCTAAGC
37	TS62 (DS5)	GCGTGGGGGCTCCGTGCCCCACGC	GGGTCCAT	*VEGFA*	3.4	VEGF_TIDE6	GAGGAAGAGTAGCTCGCCGAG	AGACCGAGTGGCAGTGACAGCAAG
38	TS63 (DS6)	GCATGGGCAGGGGCTGGGGTGCAC	AGGCCCAG	*VEGFA*	16	VEGF_TIDE7	GGGAGAGGGAAGTGTGGGGAAG	GTCTTCCTGCTCTGTGCGCACGAC
39	TS64	GAAAATTGTGATTTCCAGATCCAC	AAGCCCAA	*FANCJ*	7	FancJ_TIDE5	TTGGGGGCTCTAAGTTATGTA	CTTCATCTGTATCTTCAGGATCA
40	TS65	GAGCAGAAAAAATTGTGATTTCC	AGATCCAC	*FANCJ*	ND	FancJ_TIDE5	TTGGGGGCTCTAAGTTATGTA	CTTCATCTGTATCTTCAGGATCA

HEK293T cells were used to support transient transfections and at 72-hours post transfection the, cells were harvested, followed by genomic DNA extraction and selective amplification of the targeted locus. TIDE analysis was used to measure indel efficiency at each locus (Brinkman et al., 2014). Nme2Cas9 editing was detectable at most of these sites, even though efficiencies varied depending on the target sequence. Table 1. Interestingly, Nme2Cas9 induced indels at several genomic sites with N_4CD PAMs, albeit less consistently and at lower levels. Table 1. Fourteen (14) sites with N_4CC PAMs were analyzed in triplicate, and consistent editing was observed. See, Figure 9A. In addition, editing efficiency could be improved significantly by increasing the quantity of the Nme2Cas9 plasmid delivered, and this high efficiency could be extended to precise segmental deletion with two guides. See, Figures 10A and 10B.

The ability of Nme2Cas9 to function was tested in mouse Hepa1-6 cells (hepatoma-derived). For Hepa1-6 cells, a single plasmid encoding both Nme2Cas9 and an sgRNA (targeting either *Rosa26* or *Pcsk9*) was transiently transfected and indels were measured after 72 hrs. Editing was readily observed at both sites. See, Figure 9B, left. Nme2Cas9's functionality was also tested when stably expressed in human leukemia K562 cells. To this end, a lentiviral

construct was created expressing Nme2Cas9 and transduced cells to stably express Nme2Cas9 under the control of the SFFV promoter. This stable cell line did not show any visible differences with respect to growth and morphology in comparison to untransduced cells, suggesting that Nme2Cas9 is not toxic when stably expressed. These cells were transiently electroporated with plasmids expressing sgRNAs and analyzed by TIDE after 72 hours to measure indel efficiencies. Efficient (>50%) editing was observed at all three sites tested, validating Nme2Cas9's ability to function upon lentiviral delivery in K562 cells. See, Figure 9B.

Ribonucleoprotein (RNP) delivery of Cas9 and its sgRNA is also useful for some genome editing applications, and the greater transience of Cas9's presence can minimize off-target editing (Kim et al., 2014; Zuris et al., 2015). Moreover, some cell types (e.g. certain immune cells) are recalcitrant to DNA transfection-based editing (Schumann et al., 2015). To test whether Nme2Cas9 is functional by RNP delivery, a 6xHis-tagged Nme2Cas9 (fused to three NLSs) was cloned into a bacterial expression construct and the recombinant protein was purified. The recombinant protein was then loaded with T7 RNA polymerase-transcribed sgRNAs targeting three previously validated sites. Electroporation of the Nme2Cas9:sgRNA complex induced successful editing at each of the three target sites in HEK293T cells, as detected by TIDE. See, Figure 9C. Collectively these results indicate that Nme2Cas9 can be delivered effectively via plasmid or lentivirus, or as an RNP complex, in multiple cell types.

5. Anti-CRISPR Regulation

To date, five families of Acrs from diverse bacterial species have been shown to inhibit Nme1Cas9 in vitro and in human cells (Pawluk et al., 2016; Lee et al., 2018, submitted). Considering the high sequence identity between Nme1Cas9 and Nme2Cas9, at least some of these Acr families should inhibit Nme2Cas9. To test this, all five families of recombinant Acrs were expressed, purified and tested for Nme2Cas9's ability to cleave a target in vitro in the presence of a member of each family (10:1 Acr:Cas9 molar ratio). An inhibitor was used for the type I-E CRISPR system in *E. coli* (AcrE2) as a negative control, while Nme1Cas9was used as a positive control. (Pawluk et al., 2014); (Pawluk et al., 2016). As expected, all 5 families inhibited Nme1Cas9, while AcrE2 failed to do so. See, Figure 11A, top. AcrIIC1$_{Nme}$, AcrIIC2$_{Nme}$, AcrIIC3$_{Nme}$, and AcrIIC4$_{Hpa}$ completely inhibited Nme2Cas9. Strikingly, however, AcrIIC5$_{Smu}$ which has been previously reported as the most potent of the Nme1Cas9 inhibitors (Lee et al.,

2018), did not inhibit Nme2Cas9 in vitro even at a 10-fold molar excess. This suggests that it likely inhibits Nme1Cas9 by interacting with its PID.

Figure 10 presents exemplary data showing dose dependence and segmental deletions by Nme2Cas9, as related to Figure 9. Figure 10A shows exemplary increasing the dose of electroporated Nme2Cas9 plasmid (500 ng, vs. 200 ng in Figure 3A) improves editing efficiency at two sites (TS16 and TS6). Data provided in yellow are re-used from Figure 9A. Figure 10B shows exemplary Nme2Cas9 can be used to create precise segmental deletions. Two TLR2.0 targets with cleavage sites 32 bp apart were targeted simultaneously with Nme2Cas9. The majority of lesions created were deletions of exactly 32 bp (blue).

Figure 11 presents exemplary data showing that Nme2Cas9 is subject to inhibition by a subset of type II-C anti-CRISPR families in vitro and in cells. All experiments were done in triplicate and error bars represent s.e.m. Figure 11A shows exemplary In vitro cleavage assay of Nme1Cas9 and Nme2Cas9 in the presence of five previously characterized anti-CRISPR proteins (10:1 ratio of Acr:Cas9). Top: Nme1Cas9 efficiently cleaves a fragment containing a protospacer with an N_4GATT PAM in the absence of an Acr or in the presence of a negative control Acr (AcrE2). All five previously characterized type II-C Acr families inhibited Nme1Cas9, as expected. Bottom: Nme2Cas9 inhibition mirrors that of Nme1Cas9, except for the lack of inhibition by AcrIIC5$_{Smu}$. Figure 11B shows exemplary genome editing in the presence of the five previously described anti-CRISPR families. Plasmids expressing Nme2Cas9 (200 ng), sgRNA (100 ng) and each respective Acr (200 ng) were co-transfected into HEK293T cells, and genome editing was measured using Tracking of Indels by Decompostion (TIDE) 72 hr post transfection. Consistent with our in vitro analyses, all type II-C anti-CRISPRs except AcrIIC5$_{Smu}$ inhibited genome editing, albeit with different efficiencies. Figure 11C shows exemplary Acr inhibition of Nme2Cas9 is dose-dependent with distinct apparent potencies. Nme2Cas9 is fully inhibited by AcrIIC1$_{Nme}$ and AcrIIC4$_{Hpa}$ at 2:1 and 1:1 mass ratios of cotransfected Acr and Nme2Cas9 plasmids, respectively.

To further test this, a Nme1Cas9/Nme2Cas9 chimera with the PID of Nme2Cas9 was tested. See, Figure 5D and Figure 6D. Due to the reduced activity of this hybrid, a ~30x higher concentration of Cas9 was used to achieve a similar cleavage efficiency while maintaining the 10:1 Cas9:Acr molar ratio. No inhibition was observed by AcrIIC5$_{Smu}$ on this protein chimera. See, Figure 12. This data provides further evidence that AcrIIC5$_{Smu}$ likely interacts with the PID

of Nme1Cas9. Regardless of the mechanistic basis for the differential inhibition by AcrIIC5$_{Smu}$, these results indicate that Nme2Cas9 is subject to inhibition by the other four type II-C Acr families.

Figure 12 presents exemplary data showing that a Nme2Cas9 PID swap renders Nme1Cas9 insensitive to AcrIIC5$_{Smu}$ inhibition, as related to Figure 11. In vitro cleavage by the Nme1Cas9-Nme2Cas9PID chimera in the presence of previously characterized Acr proteins (10 uM Cas9-sgRNA + 100 uM Acr).

Based on the above in vitro data, it was hypothesized that AcrIIC1$_{Nme}$, AcrIIC2$_{Nme}$, AcrIIC3$_{Nme}$, and AcrIIC4$_{Hpa}$ could be used as off-switches for Nme2Cas9 genome editing. To test this, Nme2Cas9/sgRNA plasmid transfections (150 ng of each plasmid) targeting TS16 were performed in HEK293T cells in the presence or absence of Acr expression plasmids, as it has been reported that most Acrs inhibited Nme1Cas9 at those plasmid ratios (Pawluk et al., 2016). As expected, AcrIIC1$_{Nme}$, AcrIIC2$_{Nme}$, AcrIIC3$_{Nme}$ and AcrIIC4$_{Hpa}$ inhibited Nme2Cas9 genome editing, while AcrIIC5$_{Smu}$ had no effect. See, Figure 11B. Complete inhibition was observed by AcrIIC3$_{Nme}$ and AcrIIC4$_{Hpa}$, suggesting that they have high potency against Nme2Cas9 as compared to AcrIIC1$_{Nme}$ and AcrIIC2$_{Nme}$. To further compare the potency of AcrIIC1$_{Nme}$ and AcrIIC4$_{Hpa}$, we repeated the experiments at various ratios of Acr plasmid to Cas9 plasmid. See, Figure 11C. The data show that the AcrIIC4$_{Hpa}$ plasmid is especially potent against Nme2Cas9. Together, these data suggest that several Acr proteins can be used as off-switches for Nme2Cas9-based applications.

6. Hyper-Accuracy

Nme1Cas9 demonstrates remarkable editing fidelity in cells and mouse models (Lee et al., 2016; Amrani et al., 2018; Ibraheim et al., 2018). Furthermore, the similarity of Nme2Cas9 to Nme1Cas9 over most of its length suggests that it may likewise be hyper-accurate. However, the higher number of sites sampled in the genome as a result of the dinucleotide PAM could create more opportunities for Nme2Cas9 off-targeting in comparison with Nme1Cas9 and its less frequently encountered 4-nucleotide PAM. To assess the off-target profile of Nme2Cas9, GUIDE-seq (genome-wide, unbiased identification of double-stranded breaks enabled by sequencing) was used to identify potential off-target sites empirically and in an unbiased fashion (Tsai et al., 2014). Even the best off-target prediction algorithms are prone to false negatives necessitating empirical target site profiling methods (Bolukbasi et al., 2015b; Tsai and Joung,

2016; Tycko et al., 2016). GUIDE-seq relies on the incorporation of double-stranded oligodeoxynucleotides (dsODNs) into DNA double-stranded break sites throughout the genome. These insertion sites are then detected by amplification and high-throughput sequencing.

Because SpyCas9 is a well-characterized Cas9 ortholog it is useful for multiplexed applications with other Cas9s, and as a benchmark for their editing properties (Jiang and Doudna, 2017; Komor et al., 2017). SpyCas9 and Nme2Cas9 were cloned into identical plasmid backbones, with the same UTRs, linkers, NLSs, and promoters, for parallel transient transfections (along with similarly matched sgRNA-expressing plasmids) into HEK293T cells. First, it was confirmed that the RNA guides for SpyCas9 and Nme2Cas9 are orthogonal, i.e. that Nme2Cas9 sgRNAs do not direct editing by SpyCas9, and vice versa. See, Figure 13A. This was in contrast to earlier reported results with Nme1Cas9 (Esvelt et al., 2013; Fonfara et al., 2014).

Next, to identify a use of SpyCas9 as a benchmark for GUIDE-seq, because SpyCas9 and Nme2Cas9 have non-overlapping PAMs its can therefore potentially edit any dual site (DS) flanked by a 5'-NGGNCC-3' sequence, which simultaneously fulfills the PAM requirements of both Cas9's. This permits side-by-side comparisons of off-targeting with RNA guides that facilitate an edit of the exact same on-target site. See, Figure 14A. Six (6) DSs in *VEGFA* were targeted, each of which also has a G at the appropriate positions 5' of the PAM such that both SpyCas9 and Nme2Cas9 guides (driven by the U6 promoter) were 100% complementary to the target site. Seventy-two (72) hours after transfection, a TIDE analysis was performed on these sites targeted by each nuclease. Nme2Cas9 induced indels at all six sites, albeit at low efficiencies at two of them, while SpyCas9 induced indels at four of the six sites. See, Figure 14B. At two of the four sites (DS1 and DS4) at which SpyCas9 was effective, it induced ~7-fold more indels than Nme2Cas9, while Nme2Cas9 induced a ~3-fold higher frequency of indels than SpyCas9 at DS6. Both Cas9 orthologs edited DS2 with approximately equal efficiency.

For GUIDE-seq, DS2, DS4 and DS6 were selected to sample off-target cleavage with Nme2Cas9 guides that direct on-target editing as efficiently, less efficiently, or more efficiently than the corresponding SpyCas9 guides, respectively. In addition to the three dual sites, TS6 was added as it has been observed to be an efficiently edited Nme2Cas9 target

sites, having an approximate 30-50% indel efficiency depending on the cell type. See, Figures 9A and 10A. Similar data is seen with the mouse *Pcsk9* and *Rosa26* Nme2Cas9 sites. See, Figure 9B.

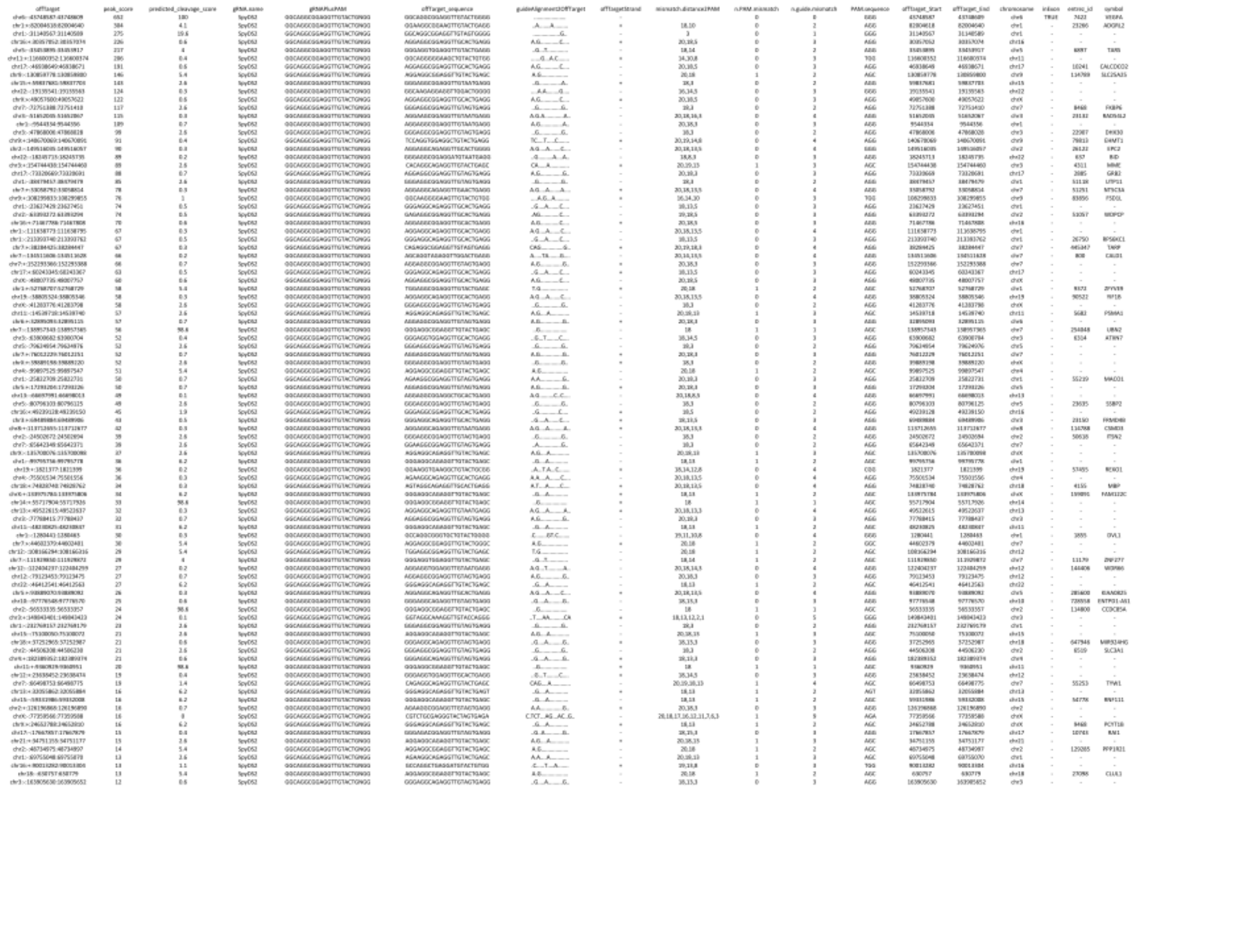
Plasmid transfections were performed for each Cas9 along with their cognate sgRNAs and the dsODNs. Subsequently, GUIDE-seq libraries were prepared as described previously (Amrani et al., 2018). A GUIDE-seq analysis revealed efficient on-target editing for both Cas9 orthologs, with relative efficiencies (as reflected by GUIDE-seq read counts) that are similar to those observed by TIDE. Figure 13B and Table 2. (Tsai et al., 2014; Zhu et al., 2017).

Figure 13 presents exemplary data showing orthogonality and relative accuracy of Nme2Cas9 and SpyCas9 at dual target sites, as related to Figure 12. Figure 13A shows exemplary Nme2Cas9 and SpyCas9 guides are orthogonal. TIDE results show the frequencies of indels created by both nucleases targeting DS2 with either their cognate sgRNAs or with the sgRNAs of the other ortholog. Figure 13B shows exemplary Nme2Cas9 and SpyCas9 exhibiting comparable on-target editing efficiencies as assessed by GUIDE-seq. Bars indicate on-target read counts from GUIDE-Seq at the three dual sites targeted by each ortholog. Orange bars represent Nme2Cas9 and black bars represent SpyCas9. Figure 13C shows an exemplary SpyCas9's on-target vs. off-target read counts for each site. Orange bars represent the on-target reads while black bars represent off-targets. Figure 13D shows exemplary Nme2Cas9's on-target vs. off-target reads for each site. Figure 13E bar graphs showing exemplary indel efficiencies (measured by TIDE) at potential off-target sites predicted by CRISPRSeek. On- and off-target site sequences are shown on the left, with the PAM region underlined and sgRNA mismatches and non-consensus PAM nucleotides given in red.

Table 2: GUIDE-seq Data

offTarget	peak_score	predicted_cl	gRNA.name	gRNAPlusPAI	offTarget_se	guideAlignmi	offTargetStra	mismatch.dis	n.PAM.mis	min.guide.mis	PAM.sequen	offTarget_Sta	offTarget_En	chromosome	inExon	entrez_id	symbol
chr6:-43748	652	100	SpyDS2	GGCAGGCGGGGCAGGCGG	.....	-			0	0	GGG	43748587	43748609	chr6	TRUE	7422	VEGFA
chr1:+82004	304	4.1	SpyDS2	GGCAGGCGGGGAAGGCGG	...A.....A.....	+	18,10		0	2	AGG	82004618	82004640	chr1		23266	ADGRL2
chr1:-31140	275	19.6	SpyDS2	GGCAGGCGGGGCAGGCGG	.....G..	-		3	0	1	GGG	31140567	31140589	chr1			
chr16:+3035	226	0.6	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....C.	+	20,18,5		0	3	AGG	30357052	30357074	chr16			
chr3:-33453	217	4	SpyDS2	GGCAGGCGGGGAGGTGG	..G...T.....	-	18,14		0	2	AGG	33453895	33453917	chr3		6897	TARS
chr11:+1166	206	0.4	SpyDS2	GGCAGGCGGGCAGGGGG	.....G...A.C....	+	14,10,8		0	3	TGG	116600352	116600374	chr11			
chr17:-4693	191	0.6	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....C.	-	20,18,5		0	3	AGG	46938649	46938671	chr17		10241	CALCOCO2
chr9:-13085	146	5.4	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....	-	20,18		1	2	AGC	130859778	130859800	chr9		114789	SLC25A25
chr15:+5983	143	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....A +		18,3		0	2	AGG	59837681	59837703	chr15			
chr22:-1913	124	0.3	SpyDS2	GGCAGGCGGGGCAAGAGG	....A.A.....G.	-	16,14,5		0	3	GGG	19135541	19135563	chr22			
chrX:+49057	122	0.6	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....C.	+	20,18,5		0	3	AGG	49057600	49057622	chrX			
chr7:-72751	117	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	72751388	72751410	chr7		8468	FKBP6
chr3:-51652	115	0.3	SpyDS2	GGCAGGCGGAGGAAGCGG	A.G.A.....	-	20,18,16,3		0	4	AGG	51652045	51652067	chr3		23132	RAD54L2
chr1:-95443	109	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....J-		20,18,3		0	3	AGG	9544334	9544356	chr1			
chr3:-47868	99	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	47868006	47868028	chr3		22907	DHX30
chr9:+14067	91	0.4	SpyDS2	GGCAGGCGGTCCAGGTGG	TC....T....C....	+	20,19,14,8		0	4	AGG	140670069	140670091	chr9		79813	EHMT1
chr2:-14951	90	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G.....A.....C-		20,18,13,5		0	4	GGG	149516035	149516057	chr2		26122	EPC2
chr22:-1824	89	0.2	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....A...J-		18,8,3		0	3	AGG	18245713	18245735	chr22		637	BID
chr3:+15474	89	2.6	SpyDS2	GGCAGGCGGCACAGGCAG	CA.....A.....	+	20,19,13		1	3	AGC	154744438	154744460	chr3		4311	MME
chr17:-7332	88	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(-		20,18,3		0	3	AGG	73320669	73320691	chr17		2885	GRB2
chr1:-38479	85	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	38479457	38479479	chr1		51118	UTP11
chr7:+33058	78	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G.....A+		20,18,13,5		0	4	AGG	33058792	33058814	chr7		51251	NT5C3A
chr9:+10829	76	1	SpyDS2	GGCAGGCGGGGCAAGGGG	....A.G...A.....+		16,14,10		0	3	TGG	108299833	108299855	chr9		83856	FSD1L
chr1:-23627	74	0.5	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....C.	-	18,13,5		0	3	AGG	23627429	23627451	chr1			
chr2:-63393	74	0.5	SpyDS2	GGCAGGCGGGAGAGGCGG	.AG.....C.	-	19,18,5		0	3	AGG	63393272	63393294	chr2		51057	WDPCP
chr16:+7146	70	0.6	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....C.	+	20,18,5		0	3	AGG	71467786	71467808	chr16			
chr1:-11163	67	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G...A.....C-		20,18,13,5		0	4	AGG	111638773	111638795	chr1			
chr1:-21339	67	0.5	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....C.	-	18,13,5		0	3	AGG	213393740	213393762	chr1		26750	RPS6KC1
chr7:+38284	67	0.3	SpyDS2	GGCAGGCGGCAGAGGCGG	CAG.....	+	20,19,18,3		0	4	AGG	38284425	38284447	chr7		445347	TARP
chr7:-13451	66	0.2	SpyDS2	GGCAGGCGGAGCAGGTAG	A.....TA.....G-		20,14,13,5		0	4	AGG	134511606	134511628	chr7		800	CALD1
chr7:+15229	66	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(+		20,18,3		0	3	AGG	152293366	152293388	chr7			
chr17:+6024	63	0.5	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....C.	+	18,13,5		0	3	AGG	60243345	60243367	chr17			
chrX:-48007	60	0.6	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....C.	-	20,18,5		0	3	AGG	48007735	48007757	chrX			
chr1:+52768	58	5.4	SpyDS2	GGCAGGCGGTGGAGGCGGT	G.....	+	20,18		1	2	AGC	52768707	52768729	chr1		9372	ZFYVE9
chr19:-3880	58	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G...A.....C-		20,18,13,5		0	4	AGG	38805324	38805346	chr19		90522	YIF1B
chrX:-41283	58	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	41283776	41283798	chrX			
chr11:-1453	57	2.6	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G.....A.....	-	20,18,13		1	3	AGC	14539718	14539740	chr11		5682	PSMA1
chr6:+32895	57	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(+		20,18,3		0	3	AGG	32895093	32895115	chr6			
chr7:-13895	56	98.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....	-		18	1	1	AGC	138957343	138957365	chr7		254048	UBN2
chr3:-63900	52	0.4	SpyDS2	GGCAGGCGGGGAGGTGG	..G...T.....C.	-	18,14,5		0	3	AGG	63900682	63900704	chr3		6314	ATXN7
chr5:-79624	52	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	79624854	79624876	chr5			
chr7:+76012	52	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(+		20,18,3		0	3	AGG	76012229	76012251	chr7			
chrX:+39889	52	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G +		18,3		0	2	AGG	39889198	39889220	chrX			
chr4:-99897	51	5.4	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....	-	20,18		1	2	AGC	99897525	99897547	chr4			
chr1:-25821	50	0.7	SpyDS2	GGCAGGCGGAGAAGGCGG	A.A.....(-		20,18,3		0	3	AGG	25822709	25822731	chr1		55219	MACO1
chr5:+17293	50	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(+		20,18,3		0	3	AGG	17293204	17293226	chr5			
chr13:-6669	49	0.1	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....C..C-		20,18,8,5		0	4	AGG	66697991	66698013	chr13			
chr5:-80796	49	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	80796103	80796125	chr5		23635	SSBP2
chr16:+4923	45	1.9	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....C.. +		18,5		0	2	AGG	49239128	49239150	chr16			
chr3:+69489	43	0.5	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....C. +		18,13,5		0	3	AGG	69489884	69489906	chr3		23150	FRMD4B
chr8:+11371	42	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G...A.....+		20,18,13,3		0	4	AGG	113712655	113712677	chr8		114788	CSMD3
chr2:-24502	39	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	24502672	24502694	chr2		50618	ITSN2
chr7:-65642	39	2.6	SpyDS2	GGCAGGCGGGGAAGGCGG	A.....G-		18,3		0	2	AGG	65642349	65642371	chr7			
chrX:-13570	37	2.6	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G.....A.....	-	20,18,13		1	3	AGC	135700076	135700098	chrX			
chr1:-99795	36	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	-	18,13		1	2	AGC	99795756	99795778	chr1			
chr19:+1821	36	0.2	SpyDS2	GGCAGGCGGGAAGGTGA	A...T.A...C....	+	18,14,12,8		0	4	CGG	1821377	1821399	chr19		57455	REXO1
chr4:-75501	36	0.3	SpyDS2	GGCAGGCGGAGAAGGCAG	A.A.....A.....C-		20,18,13,5		0	4	AGG	75501534	75501556	chr4			
chr18:+7482	34	0.3	SpyDS2	GGCAGGCGGAGTAGGCAG	A.T.....A.....C +		20,18,13,5		0	4	AGG	74828740	74828762	chr18		4155	MBP
chrX:+13397	34	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	+	18,13		1	2	AGC	133975784	133975806	chrX		159091	FAM122C
chr14:+5571	33	98.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....	+		18	1	1	AGC	55717904	55717926	chr14			
chr13:+4952	32	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G...A.....	+	20,18,13,3		0	4	AGG	49522615	49522637	chr13			
chr3:-77788	32	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(-		20,18,3		0	3	AGG	77788415	77788437	chr3			
chr11:-4823	31	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	-	18,13		1	2	AGC	48230825	48230847	chr11			
chr1:-12804	30	0.3	SpyDS2	GGCAGGCGGGCCAGGCGG	.C.....GT.C....	-	19,11,10,8		0	4	GGG	1280441	1280463	chr1		1855	DVL1
chr7:+44602	30	5.4	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....	+	20,18		1	2	GGC	44602379	44602401	chr7			
chr12:-1081	29	5.4	SpyDS2	GGCAGGCGGTGGAGGCGGT	G.....	-	20,18		1	2	AGC	108166294	108166316	chr12			
chr7:-11192	29	4	SpyDS2	GGCAGGCGGGGAGGTGG	..G...T.....	-	18,14		1	2	AGC	111929850	111929872	chr7		11179	ZNF277
chr12:-1224	27	0.2	SpyDS2	GGCAGGCGGAGGAGGTGGA	.G...T.....	-	20,18,14,3		0	4	AGG	122404237	122404259	chr12		144406	WDR66
chr12:-7912	27	0.7	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....(-		20,18,3		0	3	AGG	79123453	79123475	chr12			
chr22:-4641	27	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	-	18,13		1	2	AGC	46412541	46412563	chr22			
chr5:+93889	26	0.3	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G...A.....C+		20,18,13,5		0	4	AGG	93889070	93889092	chr5		285600	KIAA0825
chr10:-9777	25	0.6	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....(-		18,13,3		0	3	AGG	97776548	97776570	chr10		728558	ENTPD1-AS1
chr2:-56533	24	98.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....	-		18	1	1	AGC	56533335	56533357	chr2		114800	CCDC85A
chr3:+14984	24	0.1	SpyDS2	GGCAGGCGGGGTAGGCAA	..T....AA.....	+	18,13,12,2,1		0	5	GGG	149843401	149843423	chr3			
chr1:-23276	23	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	232769157	232769179	chr1			
chr15:-7510	21	2.6	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G...A.....	-	20,18,13		1	3	AGC	75100050	75100072	chr15			
chr18:+3725	21	0.6	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....(+		18,13,3		0	3	AGG	37252965	37252987	chr18		647946	MIR924HG
chr2:-44506	21	2.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....G-		18,3		0	2	AGG	44506208	44506230	chr2		6519	SLC3A1
chr4:+18238	21	0.6	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....(+		18,13,3		0	3	AGG	182389352	182389374	chr4			
chr11:+9360	20	98.6	SpyDS2	GGCAGGCGGGGAGGCGG	..G.....	+		18	1	1	AGC	9360929	9360951	chr11			
chr12:+2363	19	0.4	SpyDS2	GGCAGGCGGGGAGGTGG	..G...T.....C.	+	18,14,5		0	3	AGG	23638452	23638474	chr12			
chr7:-66498	19	1.4	SpyDS2	GGCAGGCGGCAGAGGCAG	CAG.....A.....	-	20,19,18,13		1	4	AGC	66498753	66498775	chr7		55253	TYW1
chr13:+3205	16	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	+	18,13		1	2	AGT	32055862	32055884	chr13			
chr15:-5933	16	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	-	18,13		1	2	AGC	59331986	59332008	chr15		54778	RNF111
chr2:+12619	16	0.7	SpyDS2	GGCAGGCGGAGAAGGCGG	A.A.....(+		20,18,3		0	3	AGG	126196868	126196890	chr2			
chrX:-77359	16	0	SpyDS2	GGCAGGCGGCGTCTGCGA	C.TCT...AG...J-		20,18,17,16,1		1	9	AGA	77359566	77359588	chrX			
chrX:+24652	16	6.2	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....	+	18,13		1	2	AGC	24652788	24652810	chrX		9468	PCYT1B
chr17:-1766	15	0.4	SpyDS2	GGCAGGCGGGGAGACGG	..G..A.....(-		18,15,3		0	3	AGG	17667857	17667879	chr17		10743	RAI1
chr21:+3475	15	2.6	SpyDS2	GGCAGGCGGAGGAGGCAG	A.G.....A.....	+	20,18,13		1	3	AGC	34751155	34751177	chr21			
chr2:-48734	14	5.4	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....	-	20,18		1	2	AGC	48734975	48734997	chr2		129285	PPP1R21
chr1:-69755	13	2.6	SpyDS2	GGCAGGCGGAGAAGGCAG	A.A.....A.....	-	20,18,13		1	3	AGC	69755048	69755070	chr1			
chr16:+9001	13	1.1	SpyDS2	GGCAGGCGGGCCAGGCTG	.C.....T....A....	+	19,13,8		0	3	TGG	90013282	90013304	chr16			
chr18:-6307	13	5.4	SpyDS2	GGCAGGCGGAGGAGGCGG	A.G.....	-	20,18		1	2	AGC	630757	630779	chr18		27098	CLUL1
chr3:-16390	12	0.6	SpyDS2	GGCAGGCGGGGAGGCAG	..G...A.....(-		18,13,3		0	3	AGG	163905630	163905652	chr3			

SPYDS2



offTarget	peak_score	predicted_cleavage_score	gRNA.name	gRNA.PlusPAM	offTarget_sequence	guideAlignment2OffTarget	offTargetStrand	mismatch.distanceToPAM	nPAM.mismatch	nguide.mismatch	PAM.sequence	offTarget_Start	offTarget_End	chromosome	inExon	entrez_id	symbol
chr6:-43748587-43748809	652	180	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GGCAGGGCGGAGGTTGTACTGAGG	.....	-	-	0	0	AGG	43748587	43748809	chr6	TRUE	7423	VEGFA
chr1:+:82004638:82004640	584	4.1	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GGAAGGCGGAAGTTGTACTGAGG	.A.A.....	+	18,10	0	2	AGG	82004618	82004640	chr1	-	23286	ADGRL2
chr1:-:31445567:31445589	275	19.6	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GGCAGGCGGAAGGTTGTAGTGGGG	.....G.....	-	3	0	1	GGG	31445567	31445589	chr1	-	-	-
chr16:+:8085785:8085787	226	0.6	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGCACTGAGG	.A.G.....C.	+	20,10,5	0	3	AGG	8085785	8085787	chr16	-	-	-
chr5:-:35453895:35453917	217	4	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GGGGAGGTGGAGGTTGTACTGAGG	..G.T.....	-	18,14	0	2	AGG	35453895	35453917	chr5	-	8887	TAS1
chr11:+:116680352:116680374	286	0.4	SpyD52	GGCAGGGCGGAAGCTGTACTGTGG	GGCAGGGCGGAAGCTGTACTGTGG	.....G.A.C.....	+	14,10,8	0	3	TGG	116680352	116680374	chr11	-	-	-
chr17:-:46538649:46538671	191	0.6	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGCACTGAGG	.A.G.....C.	-	20,18,5	0	3	AGG	46538649	46538671	chr17	-	10241	CALCOCO2
chr9:-:130858778:130858880	146	5.4	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGTACTGAGC	.A.G.....C.	-	20,18	1	2	AGC	130858778	130858880	chr9	-	134769	SLC25A25
chr15:+:59837800:59837802	143	2.6	SpyD52	GGCAGGGCGGAGGTTGTAATGAGG	GGGGAGGCGGAGGTTGTAATGAGG	.G.....A.	+	18,8	0	2	AGG	59837681	59837783	chr15	-	-	-
chr12:-:19139541:19139563	124	0.3	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GGCAAGAGGGAGGTTGGACTGGGG	..AA.....G.	-	16,14,5	0	3	GGG	19139541	19139563	chr12	-	-	-
chr8:+:49057600:49057622	122	0.6	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGCACTGAGG	.A.G.....C.	+	20,18,5	0	3	AGG	49057600	49057622	chr8	-	-	-
chr7:-:72751188:72751410	117	2.6	SpyD52	GGCAGGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	18,3	0	2	AGG	72751188	72751410	chr7	-	8468	PXDNL
chr8:-:91492008:91492867	116	0.3	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAAGGCGGAGGTTGTAATGAGG	.A.G.A.....A.	-	20,18,16,3	0	4	AGG	91492008	91492867	chr8	-	23132	RAD54L2
chr1:-:9944194:9944196	109	0.7	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGTAATGAGG	.A.G.....A.	-	20,18,3	0	3	AGG	9944194	9944196	chr1	-	-	-
chr3:-:47868008:47868018	99	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	10,3	0	2	AGG	47868008	47868028	chr3	-	23987	DHX30
chr9:+:140670069:140670091	91	0.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	TCCAGGTGGAGGCTGTACTGAGG	TC...T...C.....	+	20,19,14,8	0	4	AGG	140678069	140670091	chr9	-	79813	EHMT1
chr2:-:149614695:149614697	90	0.3	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCAGAGGTTGCACTGGGG	.A.G.....A.....C.	-	20,18,16,5	0	4	GGG	149614695	149616917	chr2	-	26122	EPC2
chr12:-:18245713:18245735	89	0.2	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GGGAGGGCGGAGGTTGTAATGAGG	.G.....A.....A.	-	18,8,3	0	3	AGG	18245713	18245735	chr12	-	637	BID
chr3:+:154744438:154744460	89	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	CACAGGCGGAGGTTGTACTGAGC	CA.....A.....	+	20,19,15	1	3	AGC	154744438	154744460	chr3	-	4311	MME
chr12:-:73328669:73328691	88	0.7	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTACTGAGG	.A.G.....G.	-	20,18,3	0	3	AGG	73328669	73328691	chr12	-	2885	GRK2
chr1:-:38878417:38878419	86	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	18,3	0	2	AGG	38878417	38878419	chr1	-	51138	UTP11
chr7:+:53058792:53058804	78	0.3	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCAGAGGTTGAACTGAGG	.A.G.....A.....A.	+	20,18,13,5	0	4	AGG	53058792	53058804	chr7	-	51291	MTCH2
chr9:+:108299833:108299855	76	1	SpyD52	GGCAGGGCGGAAGTTGTACTGTGG	GGCAAGGGGAAGTTGTACTGTGG	.....A.G..A.....	+	16,14,10	0	3	TGG	108299833	108299855	chr9	-	83856	PSD3L
chr1:-:23627428:23627451	74	0.5	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCAGAGGTTGCACTGAGG	.G.A.....C.....	-	10,13,5	0	3	AGG	23627428	23627451	chr1	-	-	-
chr2:-:61891272:61891294	74	0.5	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	GAAGAGGCGGAGGTTGCACTGAGG	.AG.....C.....	-	19,18,5	0	3	AGG	61891272	61891294	chr2	-	51057	MEPCE
chr16:+:71467788:71467808	70	0.8	SpyD52	GGCAGGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGCACTGAGG	.A.G.....C.	+	20,18,5	0	3	AGG	71467788	71467808	chr16	-	-	-
chr1:-:111838773:111838795	67	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGAGAGGTTGTACTGAGG	.A.G.A.....C.	-	20,18,13,5	0	4	AGG	111838773	111838795	chr1	-	-	-
chr1:-:213393740:213393762	67	0.5	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCGAGAGGTTGCACTGAGG	.G.A.....C.	-	18,13,5	0	3	AGG	213393740	213393762	chr1	-	26750	BPGM1
chr3:+:18184425:18184447	67	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	CAGAGGCGGAGGTTGTAGTGAGG	CAG.....G.....	+	20,19,18,3	0	4	AGG	18184425	18184447	chr3	-	445347	TASP
chr7:-:134511608:134511628	66	0.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGCAGGTGAGAGGTTGGACTGAGG	A...TA.....G.	-	20,16,13,5	0	4	AGG	134511608	134511628	chr7	-	808	CALD1
chr7:+:152293586:152293588	66	0.7	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGTAGTGAGG	.A.G.....G.	+	20,18,3	0	3	AGG	152293586	152293588	chr7	-	-	-
chr17:+:60243345:60243367	63	0.5	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCGAGAGGTTGCACTGAGG	.G.A.....C.	+	10,13,5	0	3	AGG	60243345	60243367	chr17	-	-	-
chr6:-:48807715:48807757	60	0.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGCACTGAGG	.A.G.....C.	-	20,18,5	0	3	AGG	48807715	48807757	chr6	-	-	-
chr1:+:52788707:52788729	58	5.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	TGGAGGCGGAGGTTGCACTGAGC	T.G.....C.....	+	20,18	1	2	AGC	52788707	52788729	chr1	-	8372	ZFYVE9
chr19:-:38805324:38805346	58	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGCACTGAGG	.A.G.A.....C.	-	20,18,13,5	0	4	AGG	38805324	38805346	chr19	-	90522	RFT1
chr6:-:41283776:41283798	58	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	10,3	0	2	AGG	41283776	41283798	chr6	-	-	-
chr11:-:14538718:14538740	57	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGAGAGGTTGTACTGAGG	.A.G.A.....C.	-	20,18,13	1	3	AGC	14538718	14538740	chr11	-	5681	PSMA1
chr8:+:128896081:128896103	57	0.7	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTAGTGAGG	.A.G.....G.	+	20,18,3	0	3	AGG	128896081	128896103	chr8	-	-	-
chr7:-:138957343:138957365	56	98.8	SpyD52	GGCAGGCGGAGGTTGTACTGAGG	GGGAGGCGGAAGTTGTACTGAGG	.G.....G.....	-	18	1	1	AGC	138957343	138957365	chr7	-	254848	UBN2
chr3:-:63808682:63808704	52	0.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGTGGAGGTTGCACTGAGG	.G.T.....C.....	-	16,14,5	0	3	AGG	63808682	63808704	chr3	-	6304	ATRN7
chr5:-:79614654:79634976	52	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	10,3	0	2	AGG	79614654	79634976	chr5	-	4645	-
chr7:+:76012229:76012251	52	0.7	SpyD52	GGCAGGCGGAGGTTGTAGTGNGG	AGGAGGGCGGAGGTTGTAGTGAGG	.A.G.....G.	+	20,18,3	0	3	AGG	76012229	76012251	chr7	-	-	-
chr8:+:59889158:59889220	52	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGNGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	+	18,3	0	2	AGG	59889198	59889220	chr8	-	-	-
chr4:-:99897525:99897547	51	5.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTACTGAGC	.A.G.....C.	-	20,18	1	2	AGC	99897525	99897547	chr4	-	-	-
chr1:-:25822709:25822731	50	0.7	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	AGAAGGCGGAGGTTGTAGTGAGG	.A.A.....G.	-	20,10,3	0	3	AGG	25822709	25822731	chr1	-	55239	MAD2L1
chr5:+:172818206:172818226	50	0.7	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTAGTGAGG	.A.G.....G.	+	20,18,3	0	3	AGG	172818204	172818226	chr5	-	-	-
chr13:-:66897981:66897803	49	0.1	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGGCGGAGGTTGCACTGAGG	.A.G.....C.C.	-	20,18,8,3	0	4	AGG	66897981	66898203	chr13	-	-	-
chr5:-:80796103:80796115	49	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	10,3	0	2	AGG	80796103	80796125	chr5	-	23635	SSBP2
chr16:+:49239128:49239150	45	1.9	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCGGAGGTTGCACTGAGG	.G.....C.	+	10,5	0	2	AGG	49239128	49239150	chr16	-	-	-
chr3:+:69489886:69489908	43	0.5	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCAGAGGTTGCACTGAGG	.G.A.....C.	+	10,13,5	0	3	AGG	69489886	69489908	chr3	-	23150	FAM43B
chr8:+:113713693:113713877	42	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGGCAGAGGTTGTAATGAGG	.A.G.....A.....A.	+	20,18,13,5	0	4	AGG	113712693	113713877	chr8	-	134788	CSMD3
chr2:-:24502672:24502894	39	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	10,3	0	2	AGG	24502672	24502894	chr2	-	50635	ITSN2
chr7:-:65642348:65642371	38	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGAAGGCGGAGGTTGTACTGAGG	.A.....G.....	-	10,3	0	2	AGG	65642348	65642371	chr7	-	-	-
chr3:-:116708076:116708098	37	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCAGAGGTTGTACTGAGC	.A.G.A.....C.	-	20,18,13	1	3	AGC	116708076	116708098	chr3	-	-	-
chr1:-:89795738:89795778	36	8.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCCAGAGGTTGTACTGAGC	.G.A.....C.....	-	18,13	1	2	AGC	89795758	89795778	chr1	-	-	-
chr19:+:3821377:3821399	36	0.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGAAGGTGAAGGCTGTACTGCGG	.A.TA.C.....	+	16,14,12,8	0	4	CGG	3821377	3821399	chr19	-	57495	RERG1
chr4:-:75505534:75505556	36	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGAAGGCAGAGGTTGCACTGAGG	.A.A.A.....C.	-	20,18,13,5	0	4	AGG	75505534	75503556	chr4	-	-	-
chr18:+:74828742:74828762	34	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGTAGGCAGAGGTTGCACTGAGG	.A.T.A.....C.	+	20,18,13,5	0	4	AGG	74828742	74828762	chr18	-	4155	MBP
chr6:+:133975783:133975806	34	8.2	SpyD52	GGCAGGCCAGAGGTTGTACTGAGC	GGCAGGCCAGAGGTTGTACTGAGC	.G.A.....C.....	+	18,13	1	2	AGC	133975783	133975806	chr6	-	139891	FAM122C
chr14:+:55717904:55717926	33	98.8	SpyD52	GGCAGGCGGAGGTTGTACTGAGC	GGGAGGCGGAGGTTGTACTGAGC	.G.....C.....	+	18	1	1	AGC	55717904	55717926	chr14	-	-	-
chr13:+:49522615:49522637	32	0.3	SpyD52	GGCAGGCGGAGGTTGTAATGAGG	AGGAGGCGGAGGTTGTAATGAGG	.A.G.....A.....A.	+	20,18,13,3	0	4	AGG	49522615	49522637	chr13	-	-	-
chr2:-:77768415:77768437	32	0.7	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	AGGAGGCGGAGGTTGTAGTGAGG	.A.G.....G.	-	20,18,3	0	3	AGG	77768415	77768437	chr2	-	-	-
chr11:-:48218801:48218807	31	8.2	SpyD52	GGCAGGCGGAGGTTGTACTGAGC	GGGAGGCCAGAGGTTGTACTGAGC	.G.A.....C.....	-	18,13	1	2	AGC	48218826	48218807	chr11	-	-	-
chr3:-:1288441:1288463	30	0.3	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GCCAGGCCGGGTGCTGTACTGGGG	.C...GT.C.....	-	19,11,10,8	0	4	GGG	1288441	1288463	chr1	-	3855	OVL3
chr7:+:44603379:44603481	29	5.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCCGGAGGTTGTACTGGGC	.A.G.....C.....	+	20,18	1	2	GGC	44603379	44603481	chr7	-	-	-
chr12:-:108166294:108166316	29	5.4	SpyD52	GGCAGGCGGAGGTTGTACTGAGC	TGGAGGCGGAGGTTGTACTGAGC	T.G.....C.....	-	20,18	1	2	AGC	108166294	108166316	chr12	-	-	-
chr7:+:111928832:111928872	28	4	SpyD52	GGCAGGCGGAGGTTGTACTGAGC	GGAGAGGTGGAGGTTGTACTGAGC	.G.T.....C.....	+	18,14	1	2	AGC	111928832	111928872	chr7	-	11179	ZNF277
chr12:-:122404237:122481259	27	2.6	SpyD52	GGCAGGCGGAGGTTGTAATGAGG	AGGAGGCTGGAGGTTGTAATGAGG	.A.G.T.....A.	-	20,18,14,5	0	4	AGG	122404237	122481259	chr12	-	146808	WDR86
chr12:-:79123453:79123475	27	0.7	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	AGGAGGCGGAGGTTGTAGTGAGG	.A.G.....G.	-	20,18,3	0	3	AGG	79123453	79123475	chr12	-	-	-
chr22:-:46412541:46412563	27	6.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCAGAGGTTGTACTGAGC	.G.A.....C.....	-	18,13	1	2	AGC	46412541	46412563	chr22	-	-	-
chr5:+:93888970:93888981	26	0.8	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCAGAGGTTGCACTGAGG	.A.G.A.....C.	+	20,18,13,5	0	4	AGG	93888970	93888981	chr5	-	285600	EANDRS
chr18:-:97778548:97778570	25	0.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.A.G.....G.	-	16,13,3	0	3	AGG	97778548	97778570	chr18	-	728558	ENTPD1-AS1
chr2:-:56533335:56533357	24	98.8	SpyD52	GGCAGGCGGAGGTTGTACTGAGC	GGGAGGCGGAGGTTGTACTGAGC	.G.....G.....	-	18	1	1	AGC	56533335	56533357	chr2	-	134800	CCDC85A
chr2:+:148843401:148843423	24	0.1	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGTAGGCAAAGGTTGTACCAGGG	.T...AA.....CA	+	18,13,12,2,1	0	5	GGG	148843401	148843423	chr2	-	-	-
chr1:-:212768137:212768179	23	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	20,3	0	2	AGG	212768137	212769179	chr1	-	-	-
chr15:-:75308050:75308073	23	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGGCAGAGGTTGTACTGAGC	.A.G.A.....C.	-	20,18,16	1	3	AGC	75308050	75308073	chr15	-	-	-
chr18:+:37252965:37252987	21	0.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.A.G.....G.	+	16,13,3	0	3	AGG	37252965	37252987	chr18	-	647948	MIR924HG
chr2:-:44506308:44506230	21	2.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCGGAGGTTGTAGTGAGG	.G.....G.	-	10,3	0	2	AGG	44506308	44506230	chr2	-	6519	SLC3A1
chr4:+:182189152:182189174	21	0.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCAGAGGTTGTAGTGAGG	.G.A.G.....G.	+	16,13,3	0	3	AGG	182189153	182189174	chr4	-	-	-
chr11:+:93800829:93800951	20	98.8	SpyD52	GGCAGGCGGAGGTTGTACTGAGC	GGGAGGCGGAGGTTGTACTGAGC	.G.....C.....	+	18	1	1	AGC	93800829	93800951	chr11	-	-	-
chr12:+:23638452:23638474	19	0.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGTGGAGGTTGTACTGAGC	.G..T.....C.....	+	16,14,5	0	3	AGC	23638452	23638474	chr12	-	-	-
chr7:-:66498753:66498775	19	1.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	CAGAGGCAGAGGTTGTACTGAGC	CAG.A.....C.....	-	20,19,16,13	1	4	AGC	66498753	66498775	chr7	-	55253	TPW1
chr11:+:32055862:32055884	16	6.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCAGAGGTTGTACTGAGT	.G.A.....C.....	+	18,13	1	2	AGT	32055862	32055884	chr11	-	-	-
chr15:-:93810986:93810008	16	8.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCCAGAGGTTGTACTGAGC	.G.A.....C.....	-	18,13	1	2	AGC	93810986	93810008	chr15	-	54756	RNF111
chr2:+:126196868:126196890	16	0.7	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	AGAAGGCGGAGGTTGTAGTGAGG	.A.A.....G.	+	20,18,3	0	3	AGG	126196868	126196890	chr2	-	-	-
chr6:-:77359566:77359588	16	8	SpyD52	GGCAGGCGGAGGTTGTACTAGTGAGA	CGTCTGCGAGGGTACTAGTGAGA	C.TCT..AG..AC.G.	-	20,18,17,16,12,11,7,6,3	1	9	AGA	77359566	77359588	chr6	-	-	-
chr9:+:24653788:24653810	16	6.2	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GGGAGGCAGAGGTTGTACTGAGC	.G.A.....C.....	+	18,13	1	2	AGC	24652788	24653810	chr6	-	8468	PCYT1B
chr17:-:17667087:17667879	15	0.4	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGAGGAGGTTGTAGTGAGG	.G.A.....G.	-	16,15,3	0	3	AGG	17667087	17667879	chr17	-	10741	RAI1
chr21:+:34791155:34751177	13	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTACTGAGC	.A.G.A.....C.	+	20,18,15	1	3	AGC	34791155	34751177	chr21	-	-	-
chr2:-:48734975:48734997	14	5.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTACTGAGC	.A.G.....C.....	-	20,18	1	2	AGC	48734975	48734997	chr2	-	129285	PPP1R21
chr1:-:60056048:60056070	13	2.6	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGAAGGCAGAGGTTGTACTGAGC	.A.A.A.....C.	-	20,18,13	1	3	AGC	60056048	60056070	chr1	-	-	-
chr16:+:80011282:80011804	13	1.1	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	GCCAGGCTGAGGATGTACTGTGG	.C...T.A.....	+	19,11,8	0	3	TGG	80011282	80011804	chr16	-	-	-
chr18:-:638757:638779	13	5.4	SpyD52	GGCAGGCGGAGGTTGTACTGNGG	AGGAGGCGGAGGTTGTACTGAGC	.A.G.....C.....	-	20,18	1	2	AGC	638757	638779	chr18	-	27088	CLUL1
chr3:-:163809630:163905652	12	0.6	SpyD52	GGCAGGCGGAGGTTGTAGTGAGG	GGGAGGCAGAGGTTGTAGTGAGG	.G.A.....G.	-	16,15,3	0	3	AGG	163905630	163905652	chr3	-	-	-

SPYDS4

offTarget	peak_score	predicted_cleavage_score	gRNA_name	gRNAPlusPAM	offTarget_sequence	guideAlignment2OffTarget	offTargetStrand	mismatch.distance2PAM	n.PAM.mismatch	n.guide.mismatch	PAM.sequence	offTarget_Start	offTarget_End	chromosome	inExon	entrez_id	symbol
chr6:-43748848:43748870	374	100	SpyDS4	GCAGGCACCTGTGCCAACATNGG	GCAGGCACCTGTGCCAACATGGG	----------	-	-	0	0	GGG	43748848	43748870	chr6	NA	7422	VEGFA
chr22:+41551021:41551043	154	0.1	SpyDS4	GCAGGCACCTGTGCCAACATNGG	ACAGGCACTGATGCCAACTTTGG	A.....TGA.....T.	+	20,12,11,10,2	0	5	TGG	41551021	41551043	chr22	TRUE	2033	EP300
chr6:+43748564:43748586	66	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	TAATGCCCTTGGAGCCTCCCTGGC	TA.T..C.TG.A...TC.C.	+	20,19,17,14,12,11,9,5,4,2	1	10	GGC	43748564	43748586	chr6	TRUE	7422	VEGFA
chrX:+77359654:77359676	47	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	GCAGGGGCGCGCCGAGAGCAGCGG	...GCG.GCC.AG.G.G	+	15,14,13,11,10,9,7,6,4,1	0	10	CGG	77359654	77359676	chrX	TRUE	5230	PGK1
chr6:-43741999:43742021	32	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	CCAGCCACCCAGCCCTCCTCCCC	C..C...CAGC..CT.C.	-	20,16,11,10,9,8,5,4,2	2	9	CCC	43741999	43742021	chr6	-	7422	VEGFA
chr15:+68132445:68132467	23	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	GTAAGCATATGATGATGTCCATTTT	.T.A...TA..ATAGTC...	+	19,17,13,12,9,8,7,6,5,4	2	10	TTT	68132445	68132467	chr15	-	-	-
chrX:-77359345:77359367	23	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	CCGCGTCCCTGCGCAAACCCAGG	C.GC.TC....C.A..CC	-	20,18,17,15,14,9,6,2,1	0	9	AGG	77359345	77359367	chrX	-	-	-
chr6:+22774978:22775000	17	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	GTGCACCCCCTGCTCCTACCCCCC	.TGCA.C....CT..T..CC	+	19,18,17,16,14,9,8,5,2,1	2	10	CCC	22774978	22775000	chr6	-	-	-
chrX:+77359596:77359618	16	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	CCAGGGAAGCAATGGCAGCGCGCC	C....G.G.AA..G..G.GC	+	20,15,13,11,10,7,4,2,1	2	9	GCC	77359596	77359618	chrX	-	-	-
chr1:+82004622:82004644	13	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	GGCGGAAGTTGTACTGAGGTGAG	.GC..A.GT...A.TG.GG.	+	19,18,15,13,12,8,6,5,3,2	1	10	GAG	82004622	82004644	chr1	-	23266	ADGRL2
chr12:-80003891:80003913	11	0	SpyDS4	GCAGGCACCTGTGCCAACATNGG	GCAGGAACTGGAGTGCACAGGTG	....A.TG.A.TGC..G	-	15,12,11,9,7,6,5,1	1	8	GTG	80003891	80003913	chr12	-	5074	PAWR

SPYDS6

offTarget	peak_score	predicted_cleavage_score	gRNA_name	gRNA/PuroPAM	offTarget_sequence	guideAlignment2OffTarget	offTargetStrand	mismatch.distance2PAM	#.PAM.mismatch	#.guide.mismatch	PAMsequence	offTarget_Start	offTarget_End	chromosome	isExon	entrez_id	symbol
chr6:+80816457:80816479	699	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	CGGCAGGGGCTGAGGGGTGCACTGG	C.....A.G...	+	20,8,5	0	3	TGG	80816457	80816479	chr6	TRUE	394	BCKDHB
chr6:-22774975:22774997	553	1.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCTAGGAGCAGGGGGTGCACTGG	....T...A.....	-	17,13,10	0	3	TGG	22774975	22774997	chr6	-	-	-
chr6:-43742023:43742045	458	100	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTGGGGTGTGCACAGG	..........A....	-	0	0	0	AGG	43742023	43742045	chr6	-	7422	VEGFA
chr7:-124498153:124498175	449	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTGAGGGGTGCACAGG	....AA...A....	-	13,12,7	0	3	AGG	124498153	124498175	chr7	-	25913	POT1
chr1:-79194107:79194129	386	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTAAGGGGTGCACAGG	....AA...A....	-	13,12,7	0	3	AGG	79194107	79194129	chr1	-	-	-
chr17:+77835740:77835762	383	5.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTAAGGGGTGCACAGG	CA.....A.....	+	20,19	0	2	AGG	77835740	77835762	chr17	-	-	-
chr19:+15313634:15313656	382	0.7	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGAAAGGGGCTGGGGTGTACACGG	...A.C....A..	+	17,12,4	0	3	GGG	15313634	15313656	chr19	-	-	-
chr12:+96610610:96610632	374	3.7	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCCGGGGCCAGGGGTGTGCACAGG	...C....A.....	+	16,10	0	2	AGG	96610610	96610632	chr12	-	2004	ELK3
chr6:-79681895:79681917	352	1.5	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGACAGGGGCCGGGGTGTGCACAGG	A.A.....C.....	-	20,18,10	0	3	AGG	79681895	79681917	chr10	-	9231	DLG5
chr6:+20250488:20250510	338	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTAGGGGTGCACCGG	....AA...A....	+	13,12,7	0	3	CGG	20250488	20250510	chr6	-	-	-
chr13:-49117085:49117105	334	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGCTGGGGTGTGCACAGG	A.....AA...A....	-	20,13,12,7	0	4	AGG	49117085	49117105	chr13	-	-	-
chr12:-80003859:80003915	330	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGCTAGGGGTGCACAGG	A.....AA...A....	-	20,13,12,7	0	4	AGG	80003893	80003915	chr12	-	5074	PAWR
chr19:+15313634:15313656	302	1.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGAAGGGGCTGGGGTGTGCACTGG	A.A..A.....	-	20,17,12	0	3	TGG	77543039	77543061	chr17	-	-	-
chr8:-85977642:85977664	299	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGCTAGGGGTGCACAGG	A.....AA...A....	-	20,13,12,7	0	4	AGG	85977642	85977664	chr8	-	-	-
chr20:-35488683:35488705	277	2.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGTGGGGGCTGGGGTGTGCACAGG	A.TG.....A.....	-	20,17,16	0	3	AGG	35488683	35488705	chr20	-	140710	SOGA1
chr11:+10027645:10027667	271	1.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGAACTGGGGTGCAGGGG	A.....AA.....	+	20,13,12	0	3	GGG	10027645	10027667	chr11	-	-	-
chr22:+38318256:38338278	268	0.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	TGGCAGGGGCTGGGGTGAACTGG	T.....A.....A..	+	20,10,3	0	3	TGG	38318256	38338278	chr22	TRUE	85377	MKCALL1
chr13:+45356854:45356876	255	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGAACTGGAGTGCAGGGG	....AA...A....	+	13,12,7	0	3	GGG	45356854	45356876	chr13	-	-	-
chr20:-31061319:31061341	231	1.9	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGCCAGGGGCTGGGGAGCAGTGG	.C.....A.....	-	18,5	0	2	AGG	31061319	31061341	chr20	-	140688	NOL4L
chr11:-69051111:69051133	229	0.7	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTGGGGGTGCACAGG	.....CTG.....	-	12,11,10	0	3	AGG	69051111	69051133	chr11	TRUE	254283	CNN2
chr17:-77637693:77637715	225	2.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	TGTGGTGGGCTGGGGTGTGCACTGG	T..GT.....A.....	-	20,17,16	0	3	TGG	77637693	77637715	chr17	-	-	-
chr11:+128772408:128772430	198	0.5	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGGAGGGGGGCTGGGGGAGCACAGG	.....AG.....	+	17,16,5	0	3	AGG	128772408	128772430	chr11	TRUE	3762	KCNJ5
chr1:-39257317:39257339	172	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGAACTGTGGAGTACACGGG	.....AA...A....	-	13,12,7,4	0	4	GGG	39257317	39257339	chr1	-	-	-
chr15:-39243269:39243291	171	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GAGCAGGGAGGCTGGGGGAGTCACTGG	.A.A...A.....	-	19,17,13,5	0	4	TGG	39243269	39243291	chr15	-	-	-
chr14:-22258408:22258430	170	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGAACTGTGAGTGCACCAG	.....AA...A....	-	13,12,7	1	3	CAG	22258408	22258430	chr14	-	-	-
chr21:+42506703:42506725	166	2.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAAGGGGCAGGGGTGCACCAG	....A.....A.....	-	15,10	1	2	CAG	42506703	42506725	chr21	-	-	-
chr7:-150036050:150036072	163	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GAGCAGGGGCAAGGGTGCACAGG	AAA.....AA.....	-	20,19,17,10,9	0	5	AGG	150036050	150036072	chr7	-	9919	RAB5B52
chr7:-1140569:1140591	162	1.5	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGCCAGGGAGCAGGGGGTGCACGGG	.C...A.A.....	-	18,13,10	0	3	GGG	1140569	1140591	chr7	-	84310	C7orf50
chr4:+40239842:40239864	154	0.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	TGGCAGCGGCTGGGGAGCACTGG	T...C.....A..	+	20,14,5	0	3	TGG	40239842	40239864	chr4	-	399	RHOH
chr22:-50743552:50743574	151	0.9	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCGCTGGGCAGGGGGTGCATCGG	...GT.....A.....	-	16,15,10	0	3	TGG	50743552	50743574	chr22	-	23654	PLXNB2
chr2:-241904500:241904522	149	3.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGTGGCTGGGGTGTCATTGG	....T.....T.....	-	14,1	0	2	TGG	241904500	241904522	chr2	-	200772	LOC200772
chr8:-135776149:135776171	146	1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGCCAGGGAGCTGGGGGTGCTCAGG	.C...A.....T..	-	18,13,2	0	3	AGG	135776149	135776171	chr8	-	7410	VAV2
chr4:+22487088:22487710	145	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	CCTCAGGGGCTGGGGTGAACAGG	CCT.....A.....	+	20,19,18,3	0	4	AGG	22487088	22487710	chr4	-	55909	BIN3
chr21:+42506703:42506725	144	4.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	TGGCAGGGGTCTGGGGTGCACAGA	T.....T.....	-	20,12	1	2	AGA	42506703	42506725	chr21	-	127002	ATP5J2
chr1:-110032844:110032866	144	4.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGCTGGGGTGTGCACAGA	A.....T.....	-	19,12,1	0	3	GGG	110032844	110032866	chr1	-	11021	ATP6V1G2
chr5:-183626625:183626647	135	0.9	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GAGCAGGGGTCTGGGGTGTGCATGGG	A.....T.....	-	19,12,8,5	0	3	GGG	183626625	183626647	chr5	-	51393	EDC3
chr20:-61928182:61928204	123	2.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GAGCAGGGGCTGGGGGGGCACTGG	A.....G.....	-	19,5	0	2	TGG	61928182	61928204	chr20	-	80279	C20ORF43
chr10:-88042752:88042774	120	0.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	TGGCAGGGGGTAAGGGGTGCACTGG	T.....TAA.....	-	20,11,10,9	0	4	TGG	88042752	88042774	chr10	-	2894	GRID1
chr17:-81311626:81311648	118	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGACAGAGGGCTGGAATGTGCACTGG	A.A..A..A.....	-	20,18,14,7	0	4	TGG	81311626	81311648	chr17	-	-	-
chr4:-1003743:1003765	117	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGCTGGAGTTCACAGG	A.....A.T..	-	20,7,4	0	3	AGG	1003743	1003765	chr4	-	-	-
chr22:+19106203:19106225	115	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGAAGGGGACCAGGGGTGCACCAG	A.A...A.CA.....	+	20,17,12,10,9	1	5	CAG	19106203	19106225	chr22	-	9993	DGCR2
chr1:+44603980:44603991	114	1.5	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGCCAGGAGAGCAGGGGTGCACAGG	.C...A.A.....	+	18,13,10	0	3	AGG	44603980	44603991	chr1	-	5792	PTPRF
chr1:+114792469:114792491	110	0.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCTGGGGCTGGGGTGTCCAGGG	.C.....CA.....	+	16,2,1	0	3	GGG	114792469	114792491	chr1	-	-	-
chr19:+38997988:38998010	110	1.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCCGGGGGCTGGGGAGCACAGG	...G.....A.....	+	16,5	0	2	AGG	38997988	38998010	chr19	-	6261	RYR1
chr2:-46897354:46897376	109	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGAGCAGGGGTGCAGAGG	A.....A.CA.....G	-	20,13,10,9,1	0	5	AGG	46897354	46897376	chr2	-	-	-
chr12:+121011672:121011694	108	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGAGGGCTGGAGTGCCCAGG	....A.....A..C.	+	14,7,2	0	3	AGG	121011672	121011694	chr12	-	9921	RNF10
chr17:-75891020:75891042	105	0.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGACAGGGGGCAGGGGGTGCCCGGG	.A.....A.....C.	-	18,10,2	0	3	GGG	75891020	75891042	chr17	-	-	-
chr9:+139220931:139220953	98	5.7	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGGAGGGGGCTCGGGGTGCACGGA	A.G.....A.....	+	20,17	1	2	GGA	139220931	139220953	chr9	TRUE	26102	DKFZP434A062
chr16:+28168625:28168647	97	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGAGCAGGGGTGCACTAGG	....AA...A...T	+	13,12,7,1	0	4	AGG	28168625	28168647	chr16	-	-	-
chr15:-74840775:74840797	92	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGAGCTGGAGTGCACAAG	A.....AA...A....	+	20,13,12,7	1	4	AAG	74840775	74840797	chr15	-	80153	EDC3
chr19:+44329443:44329465	86	0.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGAAGGCTAGGGTGTGCACAGG	....A...A.....	+	14,9,1	0	3	AGG	44329443	44329465	chr19	-	-	-
chr12:+75214528:75214550	85	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGATAGGGGGCTTGGGGGGGCACAGG	A.T.....T...G....	+	20,17,11,5	0	4	AGG	75214528	75214550	chr12	-	-	-
chr17:-46656760:46656782	82	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGAAAGCAGGGGGTGCTCAGG	....AA.A.....	-	14,11,10,2	0	4	AGG	46656760	46656782	chr17	-	80279	C20ORF43
chr16:-90077745:90077767	80	1.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGGAGGGGGTTGGGGTGCACCGG	.G.....TG.....	-	17,11,10	0	3	CGG	90077745	90077767	chr16	-	79007	DENND1
chr20:+62023611:62023633	79	2.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GAGCAGGGGCTGGGGGGGCACTGG	A.....G.....	+	19,5	0	3	TGG	62023611	62023633	chr20	-	-	-
chr12:+121013758:121013780	77	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGAGGGCTGGAGTGCCCAGG	....A.....A..C.	+	14,7,2	0	3	AGG	121013758	121013780	chr12	-	9921	RNF10
chrX:-106705923:106705945	75	1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGTGGGGGCTGGGGTGTCCCAGG	...TG.....A..C.	-	17,16,2	0	3	AGG	106705923	106705945	chrX	-	-	-
chr10:+44417540:44417562	73	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAAGGGCAGGGGTGCCCTGG	....A..A.....C.	+	15,10,2	0	3	TGG	44417540	44417562	chr10	-	283033	LOC100841
chr11:-118193407:118193429	73	1.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GAGAGGGGACTGGGGTGCAGGGG	A.AG.A.....	-	19,17,16,13	0	4	GGG	118193407	118193429	chr11	-	-	-
chr16:-13411476:13411498	73	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGACTGAGGTGCATAGG	A.....A.A.....	-	20,12,8,1	0	4	AGG	13411476	13411498	chr16	-	-	-
chr8:-8206405:8206427	73	0.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCCAGGGGCTGAGGTGCACAGG	...CA.....A.....	-	16,15,8	0	3	AGG	8206405	8206427	chr8	-	54436	SNTC1
chr16:+1517259:1517281	71	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	TGGGAGGGGGCTAGAGTGCACAGG	T..G.....AA.....	+	20,17,9,7	0	4	AGG	1517259	1517281	chr16	-	1186	CLCN7
chr1:-150849302:150849324	69	0	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	CCGCAGGGGCTGGGGTGTCCAGGG	CC.....A.....T.G	-	20,16,6,2,1	0	5	GGG	150849302	150849324	chr1	TRUE	405	AMY2
chr9:-39577711:39577733	69	1.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	CAGCAGGGGCTGGGGTGTCCCTGG	A.G.....A..TG	-	19,17,2	0	3	TGG	39577711	39577733	chr9	-	-	-
chr9:-19475308:19475330	69	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAAAGGGCCGGGGTGCCCAGG	...AA.C.....C.	-	15,14,10,2	0	4	AGG	19475308	19475330	chr9	-	-	-
chr12:-2991582:2991584	67	0.1	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCGGGGGGCTGGGGGGTGCCAGG	...AG.....G..T.	+	20,16,5,2	0	4	GGG	2991582	2991584	chr12	TRUE	83857	TMTC1
chr11:-70811278:70812300	66	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGCCCAGGGTCCACAGG	A.....CA..C....	-	20,10,9,4	0	4	AGG	70811278	70812300	chr11	-	22941	SPAG62
chr13:-89700946:89700968	62	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCTGGGGCTGGGGTGTCCAGGG	.....C.....G.	-	17,16,11	0	3	AGG	89700946	89700968	chr13	-	-	-
chr1:+110166721:110166743	60	0.6	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	AGGCAGGGGCCCGGGGTGCGCAGG	A.....C.....G.	+	20,10,2	0	3	AGG	110166721	110166743	chr1	-	271	ANP32
chr11:-114079332:114079354	58	0.2	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	CGGCAGGGGCTGGGGTGTCAGGGG	...CA.A.....T	-	15,14,12,1	0	4	CGG	114079332	114079354	chr11	-	7704	ZBTB16
chr20:-71813737:71813759	57	0.4	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCTGGGGGCTGAGGTGCCGGG	...T.....A...G.	-	16,8,2	0	3	CGG	71813737	71813759	chr20	-	55606	HCMP92
chr19:-17416518:17416540	56	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGGCAGGGGGCTGGGGGGCAAGGGG	.....G..A.....	-	20,5,1	0	3	GGG	17416518	17416540	chr19	-	-	-
chr5:-184289395:184289417	56	0.3	SpyD56	GGGCAGGGGCTGGGGTGTGCACNGG	GGACCGGGGCAGCTGGGGGGGCACAGG	.A.G.A.....G...	-	19,16,13,5	0	4	AGG	184289395	184289417	chr5	-	2049	EPHB6

SPYDS6 (con't)

Nme2DS2

Nme2DS4 Nme2DS6 ROSA26 PCSK9

For off-target identification, the analysis revealed that the DS2, DS4, and DS6 SpyCas9 sgRNAs appeared to direct editing at 93, 10, and 118 candidate off-target sites, respectively, in the normal range of off-targets when plasmid-based SpyCas9 editing is analyzed by GUIDE-seq (Fu et al., 2014; Tsai et al., 2014). In striking contrast, the DS2, DS4, and DS6 Nme2Cas9 sgRNAs appeared to direct editing at 1, 0, and 1 off-target sites, respectively. Figure 14C and Table 2. When compared to the GUIDE-seq read counts for the SpyCas9 off-targets, those of Nme2Cas9 were very low, further suggesting that Nme2Cas9 is highly specific. Figure 13C *cf.* Figure 13D. Nme2Cas9 GUIDE-seq analyses with the TS6, *Pcsk9*, and *Rosa26* yielded similar results (0, 0, and 1 off-target sites, respectively, with a modest read count for the *Rosa26*-OT1 off-target site). Figure 13C, Figure 14D, and Table 2.

Figure 14 presents exemplary data showing that Nme2Cas9 exhibits little or no detectable off-targeting in mammalian cells. Figure 14Ashows an exemplary schematic depicting dual sites (DSs) targetable by both SpyCas9 and Nme2Cas9 by virtue of their non-overlapping PAMs. The Nme2Cas9 PAM (orange) and SpyCas9 PAM (blue) are highlighted. A 24nt Nme2Cas9 guide sequence is indicated in yellow; the corresponding guide sequence for SpyCas9 would be 4nt shorter at the 5' end. Figure 14B shows an exemplary Nme2Cas9 and SpyCas9 that both induce indels at DSs. Six DSs in *VEGFA* (with $GN_3GN_{19}NGGNCC$ sequences) were selected for direct comparisons of editing by the two orthologs. Plasmids expressing each Cas9 (with the same promoter, linkers, tags and NLSs) and its cognate guide were transfected into HEK293T cells. Indel efficiencies were determined by TIDE 72 hrs post transfection. Nme2Cas9 editing was detectable at all six sites and was marginally or significantly more efficient than SpyCas9 at two sites (DS2 and DS6, respectively). SpyCas9 edited four out of the six sites (DS1, DS2, DS4 and DS6), with two sites showing significantly higher editing efficiencies than Nme2Cas9 (DS1 and DS4). DS2, DS4 and DS6 were selected for GUIDE-Seq analysis as Nme2Cas9 was equally efficient, less efficient and more efficient than SpyCas9, respectively, at these sites. Figure 14C shows exemplary Nme2Cas9 genome editing that is highly accurate in human cells. Numbers of off-target sites detected by GUIDE-Seq for each nuclease at individual target sites are shown. In addition to dual sites, we analyzed TS6 (because of its high on-target editing efficiency) and *Pcsk9* and *Rosa26* sites in mouse Hepa1-6 cells (to measure accuracy in another cell type). Figure 14D shows an exemplary targeted deep sequencing to detect indels in edited cells confirms the high Nme2Cas9 accuracy indicated by

GUIDE-seq. Figure 14E shows an exemplary sequence for the validated off-target site of the *Rosa26* guide, showing the PAM region (underlined), the consensus CC PAM dinucleotide (bold), and three mismatches in the PAM-distal portion of the spacer (red).

To validate the off-target sites detected by GUIDE-seq, a targeted deep sequencing was performed to measure indel formation at the top off-target loci following GUIDE-seq-independent editing (i.e. without co-transfection of the dsODN). While SpyCas9 showed considerable editing at most off-target sites tested and, in some instances, was more efficient than that at the corresponding on-target site, Nme2Cas9 exhibited no detectable indels at the lone DS2 and DS6 candidate off-target sites. See, Figure 14D. With the *Rosa26* sgRNA, Nme2Cas9 induced ~1% editing at the *Rosa26*-OT1 site in Hepa1-6 cells, compared to ~30% on-target editing. See, Figure 14D. It is noteworthy that this off-target site has a consensus Nme2Cas9 PAM (ACTC**CC**T) with only 3 mismatches at the PAM-distal end of the guide-complementary region (i.e. outside of the seed). See, Figure 14E. These data support and reinforce our GUIDE-seq results indicating a high degree of accuracy for Nme2Cas9 genome editing in mammalian cells.

To further corroborate the above GUIDE-Seq results, CRISPRseek was used to computationally predict potential off-target sites for two active Nme2Cas9 sgRNAs that targeted TS25 and TS47, both of which are also in *VEGFA* See, Figure 9A; (Zhu et al., 2014). Three (TS25) or four (TS47) of the most closely matched predicted sites, five with N_4CC PAMs and two with N_4CA PAMs; each had 2-5 mismatches, mostly in their PAM-distal, non-seed regions. See, Figure 13E. On- vs. off-target editing was compared after Nme2Cas9+sgRNA plasmid transfections into HEK293T cells by targeted amplification of each locus, followed by TIDE analysis. Consistently, no indels could be detected at those off-target sites for either sgRNA by TIDE, while efficient on-target editing was readily detected in DNA from the same populations of cells. Taken together, our data indicate that Nme2Cas9 is a naturally hyper-accurate genome editing platform in mammalian cells.

7. Associated Adenovirus Delivery

The compact size, small PAM, and high fidelity of Nme2Cas9 offer major advantages for in vivo genome editing using Associated Adenovirus (AAV) delivery. To test whether effective Nme2Cas9 genome editing can be achieved via single-AAV delivery, Nme2Cas9 was cloned with its sgRNA and their promoters (U1a and U6, respectively) into an AAV

vector backbone. See, Figure 15A. An all-in-one AAV was prepared with an sgRN-.Nme2Cas9 packaged into a hepatotropic AAV8 capsid to target two genes in the mouse liver: i) *Rosa26* (a commonly used safe harbor locus for transgene insertion) (Friedrich and Soriano, 1991) as a negative control; and ii) *Pcsk9*, a major regulator of circulating cholesterol homeostasis (Rashid et al., 2005), as a phenotypic target.

SauCas9- or Nme1Cas9-induced indels in *Pcsk9* in the mouse liver results and reduced cholesterol levels providing a useful and easy-to-score in vivo benchmark for new editing platforms (Ran et al., 2015; Ibraheim et al., 2018). The Nme2Cas9 RNA guides were the same as those used above. See, Figure 9B, Figure 13D, and Figure 14. As *Rosa26*-OT1 was the only Nme2Cas9 off-target site that has been validated in cultured mammalian cells, the *Rosa26* guide also provided us with an opportunity to assess on- vs. off-target editing in vivo. See, Figures 14D-E). The tail veins of two groups of mice ($n = 5$) were injected with 4×10^{11} AAV8.sgRNA.Nme2Cas9 genome copies (GCs) targeting either *Pcsk9* or *Rosa26.* Serum was collected at 0, 14 and 28 days post-injection for cholesterol level measurement. Mice were sacrificed at 28 days post-injection and liver tissues were harvested. See, Figure 15A. Targeted deep sequencing of each locus revealed ~38% and ~46% indel induction at the *Pcsk9* and *Rosa26* editing sites, respectively, in the liver. See, Figure 15B. Because hepatocytes constitute only 65-70% of total cellular content in the adult liver, Nme2Cas9 AAV-induced hepatocyte editing efficiencies with sg*Pcsk9* and sg*Rosa* were approximately 54-58% and 66-71%, respectively (Racanelli and Rehermann, 2006).

Only 2.25% liver indels overall (~3-3.5% in hepatocytes) were detected at the *Rosa26*-OT1 off-target site, comparable to the 1% editing that we observed at this site in transfected Hepa1-6 cells. Figure 15B *cf* Figure 14D. At both 14 and 28 days post-injection, *Pcsk9* editing was accompanied by a ~44% reduction in serum cholesterol levels, whereas mice treated with the sg*Rosa26*-expressing AAV maintained normal level of cholesterol throughout the study. See, Figure 15C. The ~44% reduction in serum cholesterol in the Nme2Cas9/sg*Pcsk*9 AAV-treated mice compares well with the ~40% reduction reported with SauCas9 all-in-one AAV when targeting the same gene (Ran et al., 2015).

Figure 15 presents exemplary data showing Nme2Cas9 genome editing in vivo via all-in-one AAV delivery. Figure 15A shows exemplary workflow for delivery of AAV8.sgRNA.Nme2Cas9 to lower cholesterol levels in mice by targeting *Pcsk9*. Top: schematic

of the all-in-one AAV vector expressing Nme2Cas9 and the sgRNA (individual genome elements not to scale). BGH, bovine growth hormone poly(A) site; HA, epitope tag; NLS, nuclear localization sequence; h, human-codon-optimized. Bottom: Timeline for AAV8.sgRNA.Nme2Cas9 tail-vein injections (4 x 10^{11} GCs), followed by cholesterol measurements at day 14 and indel, histology and cholesterol analyses at day 28 post-injection. Figure 15B shows an exemplary TIDE analysis to measure indels in DNA extracted from livers of mice injected with AAV8.Nme2Cas9+sgRNA targeting *Pcsk9* and *Rosa26* (control) loci. Indel efficiency at the lone off-target site identified by GUIDE-seq for these two sgRNAs (Rosa26|OT1) were also assessed by TIDE. Figure 15C shows an exemplary reduced serum cholesterol levels in mice injected with the *Pcsk9*-targeting guide compared to the *Rosa26*-targeting controls. *P* values are calculated by unpaired two-tailed t-test. Figure 16 presents exemplary data showing PCSK9 knockdown and liver histology following Nme2Cas9 AAV delivery and editing, related to Figure 15. Figure 16A shows exemplary Western blotting using anti-PCSK9 antibody reveals strongly reduced levels of PCSK9 in the livers of mice treated with sg*Pcsk9*, compared to mice treated with sg*Rosa26.* 2ng of recombinant PCSK9 was used as a mobility standard (left-most lane), and a cross-reacting band in the liver samples is indicated by an asterisk. GAPDH was used as loading control (bottom panel). Figure 16B shows exemplary H&E staining from livers of mice injected with AAV8.Nme2Cas9+sg*Rosa26* (left) or AAV8.Nme2Cas9+sg*Pcsk9* (right) vectors. Scale bars, 25 μm.

Western blotting was performed using an anti-PCSK9 antibody to estimate PCSK9 protein levels in the livers of mice treated with sg*Pcsk9* and sg*Rosa26*. Liver PCSK9 was below the detection limit in mice treated with sg*Pcsk9*, whereas sg*Rosa26*-treated mice exhibited normal levels of PCSK9. See, Figure 16A. Hematoxylin and eosin (H&E) staining and histology revealed no signs of toxicity or tissue damage in either group after Nme2Cas9 expression. See, Figure 16B. These data validate Nme2Cas9 as a highly effective genome editing system in vivo, including when delivered by single-AAV vectors.

AAV vectors have recently been used for the generation of genome-edited mice, without the need for microinjection or electroporation, simply by soaking the zygotes in culture medium containing AAV vector(s), followed by reimplantation into pseudopregnant females (Yoon et al., 2018). Editing was obtained previously with a dual-AAV system in which SpyCas9 and its sgRNA were delivered in separate vectors (Yoon et al., 2018). To test

whether Nme2Cas9 could perform accurate and efficient editing in mouse zygotes with an all-in-one AAV delivery system, we targeted *Tyrosinase* (*Tyr*). A bi-allelic inactivation of *Tyr* disrupts melanin production resulting in an albino phenotype (Yokoyama et al., 1990).

An efficient *Tyr* sgRNA was validated that cleaves the *Tyr* locus only seventeen (17) bp from the site of the classic *albino* mutation in Hepa1-6 cells by transient transfections. See, Figure 17A. Next, C57BL/6NJ zygotes were incubated for 5-6 hours in culture medium containing $3x10^9$ or $3x10^8$ GCs of an all-in-one AAV6 vector expressing Nme2Cas9 along with the *Tyr* sgRNA. After overnight culture in fresh media, those zygotes that advanced to the two-cell stage were transferred to the oviduct of pseudopregnant recipients and allowed to develop to term. See, Figure 18A. Coat color analysis of pups revealed mice that were albino, chinchilla (indicating a hypomorphic allele of *Tyrosinase*), or that had variegated coat color composed of albino and chinchilla spots but lacking black pigmentation. See, Figures 18B-C. These results suggest a high frequency of biallelic mutations since the presence of a wild-type *Tyrosinase* allele should render black pigmentation. A total of five pups (10%) were born from the $3x10^9$ GCs experiment. All of them carried indels; phenotypically, two were albino, one was chinchilla, and two had variegated pigmentation, indicating mosaicism.

From the $3x10^8$ GCs experiment, four (4) pups (14%) were obtained, two of which died at birth, preventing a coat color or genome analysis. Coat color analysis of the remaining two pups revealed one chinchilla and one mosaic pup. These results indicate that single-AAV delivery of Nme2Cas9 and its guide can be used to generate mutations in mouse zygotes without microinjection or electroporation.

To measure on-target indel formation in the *Tyr* gene, DNA was isolated from the tails of each mouse, the locus was amplified and upon which a TIDE analysis was performed. All mice had high levels of on-target editing by Nme2Cas9, varying from 84% to 100%. See, Figures 17B-C. Most lesions in albino mouse 9-1 were either a 1- or a 4-bp deletion, suggesting either mosaicism or trans-heterozygosity, but albino mouse 9-2 exhibited a uniform 2-bp deletion. See, Figure 17C. Figure 17 presents exemplary data showing *Tyr* editing ex vivo in mouse zygotes, related to Figure 16. Figure 17A shows an exemplary two sites in *Tyr*, each with N_4CC PAMs, were tested for editing in Hepa1-6 cells. The sg*Tyr2* guide exhibited higher editing efficiency and was selected for further testing. Figure 17B shows an exemplary seven mice that survived post-natal development, and each exhibited coat color phenotypes as

well as on-target editing, as assayed by TIDE. Figure 17C shows an exemplary Indel spectra from tail DNA of each mouse from (B), as well as an unedited C57BL/6NJ mouse, as indicated by TIDE analysis. Efficiencies of insertions (positive) and deletions (negative) of various sizes are indicated.

Figure 18 presents exemplary data showing Nme2Cas9 genome editing ex vivo via all-in-one AAV delivery. Figure 18A shows an exemplary workflow for single-AAV Nme2Cas9 editing ex vivo to generate albino C57BL/6NJ mice by targeting the *Tyr* gene. Zygotes are cultured in KSOM containing AAV6.Nme2Cas9:sg*Tyr* for 5-6 hours, rinsed in M2, and cultured for a day before being transferred to the oviduct of pseudo-pregnant recipients. Figure 18B shows exemplary albino (left) and chinchilla or variegated (middle) mice generated by $3x10^9$ GCs, and chinchilla or variegated mice (right) generated by $3x10^8$ GCs of zygotes with AAV6.Nme2Cas9:sg*Tyr*. Figure 18C shows an exemplary summary of Nme2Cas9.sg*Tyr* single-AAV ex vivo *Tyr* editing experiments at two AAV doses.

The data is inconclusive as to whether there was no mosaicism in mouse 9-2, or that additional alleles were absent from mouse 9-1, because only tail samples were sequenced and other tissues could have distinct lesions. Analysis of tail DNA from chinchilla mice revealed the presence of in-frame mutations that are potentially the cause of the chinchilla coat color. The limited mutational complexity suggests that editing occurred early during embryonic development in these mice. These results provide a streamlined route toward mammalian mutagenesis through the application of a single AAV vector, in this case delivering both Nme2Cas9 and its sgRNA.

Figure 19 shows an exemplary mCherry reporter assay for nSpCas9-ABEmax and optimized nNme2Cas9-ABEmax activities. Figure 19A shows exemplary sequence information of ABE-mCherry reporter. There is a TAG stop codon in the mCherry coding region. In the reporter-integrated stable cell line, there is no mCherry signal due to this stop codon. The mCherry signal will be activated if the nSpCas9-ABEmax or optimized nNme2Cas9-ABEmax can convert TAG to CAG, which encodes a glutamine residue. Figure 19B shows an exemplary mCherry signal is activated due to SpCas9-ABE or Nme2Cas9-ABE activity. Upper panel: negative control (no editing); middle panel: mCherry activation by nSpCas9-ABEmax; bottom panel: mCherry activation by optimized nNme2Cas9-ABEmax. Figure 19C shows an exemplary FACS quantitation of base editing events in mCherry reporter cells transfected with the SpCas9-

ABE or Nme2Cas9-ABE. N = 6; error bars represent S.D. Results are from three biological replicates performed in technical duplicates.

Figure 20 shows an exemplary GFP reporter assay for nSpCas9-CBE4 (Addgene #100802) and nNme2Cas9-CBE4 (same plasmid backbone as Addgene #100802) activities. Figure 20A shows exemplary sequence information of the CBE-GFP reporter. There is a mutation that converts GYG to GHG in the fluorophore core region of the GFP reporter line. There is no GFP signal due to this mutation. The GFP signal will be activated if the nSpCas9-CBE4 or nNme2Cas9-CBE4 can convert CAC (encoding histidine) to TAC/TAT (encoding tyrosine). Figure 20B shows an exemplary GFP signal is activated due to nSpCas9-CBE4 or nNme2Cas9-CBE4 activity. Upper panel: negative control (no editing); middle panel: GFP activation by nSpCas9-CBE4; bottom panel: GFP activation by nNme2Cas9-CBE4). Figure 20C shows an exemplary FACS quantitation of base editing events in GFP reporter cells transfected with nSpCas9-CBE4 or nNme2Cas9-CBE4. N = 6; error bars represent S.D. Results are from biological replicates performed in technical duplicates.

Figure 21 shows exemplary cytosine editing by nNme2Cas9-CBE4. Upper panel shows the *KANK3* targeting sequence information (PAM sequences are indicated in red) of Nme2Cas9 and base editing in the negative control samples. Bottom panel shows the quantification of the substitution efficiency of each type of base in the nNmeCas9-CBE4 editing window of the *KANK3* target sequences. Sequence tables show nucleotide frequencies at each position. Frequencies of expected C-to-T conversion are indicated in red.

Figure 22 shows exemplary cytosine and adenine editing by nNme2Cas9-CBE4 and nNme2Cas9-ABEmax, respectively. Upper panel shows the *PLXNB2* targeting sequence information (PAM sequences are indicated in red) of Nme2Cas9 and base editing in the negative control samples. Middle panel shows the quantification of the substitution rate of each type of base in the nNmeCas9-ABEmax editing windows of the *PLXNB2* target sequence. Sequence tables show nucleotide frequencies at each position. Frequencies of expected A-to-G conversion are highlighted in red. Bottom panel shows the quantification of the substitution efficiency of each type of base in the nNmeCas9-CBE4 editing windows of the *PLXNB2* target sequence. Sequence tables show nucleotide frequencies at each position. Frequencies of expected C-to-T conversion are highlighted in red.

8. Sequences

Alignment of Nme1Cas9 and Nme2Cas9

Non-PID aa differences (teal- underlined); PID aa differences (yellow - underlined bold); active site residues (red **- bold).**

Nme1Cas9 (1-60)
MAAFKPNSINYILGL**D**IGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAM
Nme2Cas9 (1-60)
MAAFKPNPINYILGL**D**IGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAM

Nme1Cas9 (61-120)
ARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAAL
DR
Nme2Cas9 (61-120)
ARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAAL
DR

Nme1Cas9 (121-180)
KLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPA
EL
Nme2Cas9 (121-180)
KLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPA
EL

Nme1Cas9 (181-240)
ALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLM
Nme2Cas9 (181-240)
ALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLM

Nme1Cas9 (241-300)
TQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTD
T

Nme2Cas9 (241-300)
TQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTD
T

Nme1Cas9 (301-360)
ERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISR
AL
Nme2Cas9 (301-360)
ERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISR
AL

Nme1Cas9 (361-420)
EKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKF
Nme2Cas9 (361-420)
EKEGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKF

Nme1Cas9 (421-480)
VQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRA
Nme2Cas9 (421-480)
VQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRA

Nme1Cas9 (481-540)
LSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFRE
Y
Nme2Cas9 (481-540)
LSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFRE
Y

Nme1Cas9 (541-600)
FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEID**H**ALPFSRTWDDSF

Nme2Cas9 (541-600)

FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEID**H**ALPFSRTWDDSF

Nme1Cas9 (601-660)

NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFD
ED

Nme2Cas9 (601-660)

NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFD
ED

Nme1Cas9 (661-720)

GFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRA
END

Nme2Cas9 (661-720)

GFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAE
ND

Nme1Cas9 (721-780)

RHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWE
FFA

Nme2Cas9 (721-780)

RHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWE
FFA

Nme1Cas9 (781-840)

QEVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMS
G

Nme2Cas9 (781-840)

QEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMS
G

Nme1Cas9 (841-895)

QGHMET**VK**SAK**---RLD**E**GV**SV**L**RV**P**LT**QL**KL**K**DLE**K**MVN**R--**

ERE**PK**LYEALKARLEA**H**

Nme2Cas9 (841-899)

AHK-

DT**LR**SAK**RFVKHN**E**KI**SV**K**RV**W**LT**EI**KL**A**DLE**N**MVN**YKNG**RE**IE**LYEALKARLEA**Y**

Nme1Cas9 (896-950)

KDDPAKAF**AE---**PFYK**YDKA**G**NRT**Q**Q**VKAVRVE**QV**Q**KT**GV**WVRNH--**

NGIADN**AT**MVRV

Nme2Cas9 (900 –954)

GGNAKQAF**DPKDN**PFYK**K---**G**--**

GQ**L**VKAVRVE**KT**Q**ES**GV**LLNKKNAYT**IADN**GD**MVRV

Nme1Cas9 (951-1005)

DVF**E**K**G-----**

DKY**YL**VPIY**S**WQVA**KG**ILPD**RAVVQGKDEEDWQL**IDDS**FN**F**K**FSLH**PN**D

Nme2Cas9 (955-1007)

DVF**C**K**VDKKGKNQ**Y**FI**VPIY**A**WQVA**EN**ILPD**IDCKG-------YR**IDDS**YT**F**C**FSLH**KY**D

Nme1Cas9 (1006-1063)

L**VEVIT--**

K**KARM**F**G**Y**FAS**C**HRGT**G**NINIRI**HD**LDH**K**IGKNGILEG**I**GVKTALSF**QKYQ**I**

Nme2Cas9 (1008-1063)

L**IAFQKDE**K**SKVE**F**A**Y**YIN**C**DSSN**G**RFYLAW**HD**KGS**K**EQ----**

QFRI**STQNLVLI**QKYQ**V**

Nme1Cas9 (1064-1082)

DELGKEIRPCRLKKRPPVR

Nme2Cas9 (1064-1082)

NELGKEIRPCRLKKRPPVR

Alignment of Nme1Cas9 and Nme3Cas9

Non-PID aa differences (teal- underlined); PID aa differences (yellow - underlined bold); active site residues (red **- bold).**

Nme1Cas9 1 MAAFKPNSINYILGL**D**IGIASVGWAMVEIDEEENPIRLIDLGVRVFERAE 50
Nme3Cas9 1 MAAFKPNPINYILGL**D**IGIASVGWAMVEIDEEENPIRLIDLGVRVFERAE 50

Nme1Cas9 51
VPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDEN 100
Nme3Cas9 51
VPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDEN 100

Nme1Cas9 101
GLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGET 150
Nme3Cas9 101
GLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGET 150

Nme1Cas9 151
ADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYS 200
Nme3Cas9 151
ADKELGALLKGVADNAHALQTGDFRTPAELALNKFEKECGHIRNQRGDYS 200

Nme1Cas9 201 HTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA 250
Nme3Cas9 201
HTFSRKDLQAELNLLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA 250

Nme1Cas9 251
VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDT 300

Nme3Cas9 251
VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDT 300

Nme1Cas9 301
ERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEM 350
Nme3Cas9 301
ERATLMDEPYRKSKLTYAQARKLLSLEDTAFFKGLRYGKDNAEASTLMEM 350

Nme1Cas9 351 KAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLK 400
Nme3Cas9 351 KAYHTISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLK 400

Nme1Cas9 401 DRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYG 450
Nme3Cas9 401 DRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYG 450

Nme1Cas9 451 DHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPAR 500
Nme3Cas9 451 DHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPAR 500

Nme1Cas9 501 IHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKS 550
Nme3Cas9 501 IHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKS 550

Nme1Cas9 551
KDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEID**H**ALPFSRTWDDSF 600

Nme3Cas9 551

KDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEID**H**ALPFSRTWDDSF 600

Nme1Cas9 601

NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQ 650

Nme3Cas9 601

NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQ 650

Nme1Cas9 651

RILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNG 700

Nme3Cas9 651

RILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNG 700

Nme1Cas9 701

QITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEM 750

Nme3Cas9 701

QITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEM 750

Nme1Cas9 751

NAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEA 800

Nme3Cas9 751

NAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEA 800

Nme1Cas9 801

DTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSA 850

Nme3Cas9 801

DTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSA 850

Nme1Cas9 851

KRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPA 900

Nme3Cas9 851
KRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPA 900

Nme1Cas9
901 KAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRV 950
Nme3Cas9
901 KAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRV 950

Nme1Cas9 951
DVFEKGDKYYLVPIYSWQVAKGILPDRAVV**QGK**DEEDW**QL**ID**D**SF**N**FKF**S** 1000
Nme3Cas9 951
DVFEKGDKYYLVPIYSWQVAKGILPDRAVV**AYA**DEEDW**TV**ID**E**SF**R**FKF**V** 1000

Nme1Cas9 1001
L**HP**NDL**VE**V**IT**KK**ARMF**GYF**ASCH**R**G**TG**N**I**NI**R**I**HDL**DHKI**GK**N**G**ILEG**I 1050
Nme3Cas9 1001 L**YS**NDL**IK**V**QL**KK**DSFL**GYF**SGLD**R**A**TG**A**I**SL**R**E**HDL**EKSK**GK**D**G-**MHR**I 1049

Nme1Cas9 1051 GVKTALSFQKYQIDE**L**GKEIRPCRLKKRPPVR 1082
Nme3Cas9 1050 GVKTALSFQKYQIDE**M**GKEIRPCRLKKRPPVR 1081

Plasmid-Expressed Nme2Cas9

SV40 NLS (yellow- **BOLD**); 3X-HA-Tag (green-(**underlined/bold**); cMyc-like NLS (teal-plain); Linker (magenta - ***bold italics***) and *Nme2Cas9* (*italics*).

MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRL
ARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEW
SAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKES
GHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAV
QKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSK
LTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSE

LQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRY
DEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETA
REVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLY
SGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSR
EWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGK
RRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFD
GKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSS
RPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVN
YKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKN
AYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFCFSLH
KYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQKYQVNELG
*KEIRPCRLKKRPPVR**GTGG***<u>**YPYDVPDYAGYPYDVPDYAGSYPYDVPDYA**</u>***G***
SAAPAAKKKKLDFESG*

<u>*AAV-expressed Nme2Cas9*</u>

SV40 NLS (yellow- **BOLD**); 3X-HA-Tag (green-(<u>**underlined/bold**</u>); Nucleoplasmin-like NLS (red-<u>underline</u>); c-myc NLS (teal- plain); Linker (magenta - ***bold italics***) and *Nme2Cas9* (*italics*).

*MV***PKKKRKV*****ED***<u>KRPAATKKAGQAKKKK</u>*MAAFKPNPINYILGLDIGIASVGWAMVEIDEE*
ENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAAD
FDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALL
KGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKE
FGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKL
NNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEAST
LMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEA
LLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNP
VVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFRE
YFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNN
KVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFK
ECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHAL
DAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIR

VFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSA
KRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKD
NPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPI
YAWQVAENILPDIDCKGYRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLA
*WHDKGSKEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR**ED***KRPAATKKAGQAK
KKK**YPYDVPDYAGYPYDVPDYAGSYPYDVPDYA*****AA***PAAKKKKLD*

Recombinant Nme2Cas9

SV40 NLS (yellow- **BOLD**); Nucleoplasmin-like NLS (red-underline); Linker (magenta - ***bold italics***) and *Nme2Cas9* (*italics*).

PKKKRKV***NA***
MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKT
GDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAA
LDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAE
LALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQ
RPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATL
MDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKD
KKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVP
LMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGS
PARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQ
QHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYF
NGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHI
LLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRY
KEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRT
LLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLA
DLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQES
GVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDD
SYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQ
*KYQVNELGKEIRPCRLKKRPPVR**GGGGSGGGGSGGGGS***
PAAKKKKLD***GGGS***KRPAATK
KAGQAKKKK*

Recombinant Nme2Cas9 for use in mammalian cell RNP delivery:

SV40 NLS (yellow- **BOLD**); Nucleoplasmin-like NLS (red-<u>underline</u>); Linker (magenta - ***bold italics***) and *Nme2Cas9* (*italics*).

PKKKRKV*NA**MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKT*
GDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAA
LDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAE
LALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQ
RPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATL
MDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKD
KKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVP
LMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGS
PARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQ
QHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYF
NGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHI
LLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRY
KEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRT
LLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLA
DLENMVNYKNGREIELYEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQES
GVLLNKKNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDD
SYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRISTQNLVLIQ
*KYQVNELGKEIRPCRLKKRPPVR**GGGGSGGGGSGGGGS*****PAAKKKKLD*GGGS*<u>*KRPAATKK*</u>
<u>*AGQAKKKK*</u>*

9. Therapeutic Applications

Although compact Cas9 orthologs have been previously validated for genome editing, including via single-AAV delivery, their longer PAMs have restricted therapeutic development due to target site frequencies that are lower than that of the more widely adopted SpyCas9. In addition, SauCas9 and its KKH variant with relaxed PAM requirements (Kleinstiver et al., 2015) are prone to off-target editing with some sgRNAs (Friedland et al., 2015; Kleinstiver et al., 2015). These limitations are exacerbated with target loci that require editing within a narrow sequence window, or that require precise segmental deletion. We have identified Nme2Cas9 as a compact and highly accurate Cas9 with a less restrictive dinucleotide PAM for genome editing by AAV delivery in vivo. The development of Nme2Cas9 greatly expands the genomic scope of

in vivo editing, especially via viral vector delivery. The Nme2Cas9 all-in-one AAV delivery platform established in this study can in principle be used to target as wide a range of sites as SpyCas9 (due to the identical densities of optimal N_4CC and NGG PAMs), but without the need to deliver two separate vectors to the same target cells. The availability of a catalytically dead version of Nme2Cas9 (dNme2Cas9) also promises to expand the scope of applications such as CRISPRi, CRISPRa, base editing, and related approaches (Dominguez et al., 2016; Komor et al., 2017). Moreover, Nme2Cas9's hyper-accuracy enables precise editing of target genes, potentially ameliorating safety issues resulting from off-target activities. Perhaps counterintuitively, the higher target site density of Nme2Cas9 (compared to that of Nme1Cas9) does not lead to a relative increase in off-target editing for the former. Similar results have been reported recently with SpyCas9 variants evolved to have shorter PAMs (Hu et al., 2018). Type II-C Cas9 orthologs are generally slower nucleases in vitro than SpyCas9 (Ma et al., 2015; Mir et al., 2018); interestingly, enzymological principles indicate that a reduced apparent k_{cat} (within limits) can improve on- vs. off-target specificity for RNA-guided nucleases (Bisaria et al., 2017).

The discovery of Nme2Cas9 and Nme3Cas9 hinged on unexplored Cas9s that are highly related (outside of the PID) to an ortholog that was previously validated for human genome editing (Esvelt et al., 2013; Hou et al., 2013; Lee et al., 2016; Amrani et al., 2018). The relatedness of Nme2Cas9 and Nme3Cas9 to Nme1Cas9 brought an added benefit, namely that they use the exact same sgRNA scaffold, circumventing the need to identify and validate functional tracrRNA sequences for each. In the context of natural CRISPR immunity, the accelerated evolution of novel PAM specificities could reflect selective pressure to restore targeting of phages and MGEs that have escaped interference through PAM mutations (Deveau et al., 2008; Paez-Espino et al., 2015). Our observation that AcrIIC5$_{Smu}$ inhibits Nme1Cas9 but not Nme2Cas9 suggests a second, non-mutually-exclusive basis for accelerated PID variation, namely evasion of anti-CRISPR inhibition. We also speculate that accelerated variability may not be restricted to PIDs, perhaps resulting from selective pressures to evade anti-CRISPRs that bind other Cas9 domains. Cas9 inhibitors such as AcrIIC1 that bind more conserved regions of Cas9 likely present fewer routes toward mutational escape and therefore exhibit a broader inhibitory spectrum (Harrington et al., 2017a). Whatever the sources of selective pressure driving Acr and Cas9 co-evolution, the availability of validated inhibitors of Nme2Cas9 (e.g. AcrIIC1-4) provides opportunities for additional levels of control over its activities.

The approach used in this study (i.e. searching for rapidly-evolving domains within Cas9) can be implemented elsewhere, especially with bacterial species that are well-sampled at the level of genome sequence. This approach could also be applied to other CRISPR-Cas effector proteins such as Cas12 and Cas13 that have also been developed for genome or transcriptome engineering and other applications. This strategy could be especially compelling with Cas proteins that are closely related to orthologs with proven efficacy in heterologous contexts (e.g. in eukaryotic cells), as was the case for Nme1Cas9. The application of this approach to meningococcal Cas9 orthologs yielded a new genome editing platform, Nme2Cas9, with a unique combination of characteristics (compact size, dinucleotide PAM, hyper-accuracy, single-AAV deliverability, and Acr susceptibility) that promise to accelerate the development of genome editing tools for both general and therapeutic applications.

Table 3. The following presents exemplary sequences for plasmids and oligos as disclosed herein.

Exemplary Plasmids					
Plasmid #	**Name**	**Insert description**	**Backbone**	**Purpose**	**Insert Sequence**
1	pAE70	Nme3Cas9 PID on Nme1Cas9	pMCSG7	Bacterial expression of Nme1Cas9 with Nme3Cas9 PID	See examples herein.
2	pAE71	Nme2Cas9 PID on Nme1Cas9	pMCSG7	Bacterial expression of Nme1Cas9 with Nme2Cas9 PID	See examples herein.
3	pAE113	Nme2TLR1	pLKO	Targeting TLR2.0 with Nme2Cas9	GTCACCTGCCTCGTGGAATACGG
4	pAE114	Nme2TLR2	pLKO	Targeting TLR2.0 with Nme2Cas9	GCACCTGCCTCGTGGAATACGGT
5	pAE115	Nme2TLR5	pLKO	Targeting TLR2.0 with Nme2Cas9	GTTCAGCGTGTCCGGCTTTGGC
6	pAE116	Nme2TLR11	pLKO	Targeting TLR2.0	GTGGTGAGCAAGG

				with Nme2Cas9	GCGAGGAGCTG
7	pAE117	Nme2TLR12	pLKO	Targeting TLR2.0 with Nme2Cas9	GGGCGAGGAGCTG TTCACCGGGGT
8	pAE118	Nme2TLR13	pLKO	Targeting TLR2.0 with Nme2Cas9	GTGAACTTGTGGCC GTTTACGTCG
9	pAE119	Nme2TLR14	pLKO	Targeting TLR2.0 with Nme2Cas9	GCGTCCAGCTCGAC CAGGATGGGC
10	pAE120	Nme2TLR15	pLKO	Targeting TLR2.0 with Nme2Cas9	GCGGTGAACAGCT CCTCGCCCTTG
11	pAE121	Nme2TLR16	pLKO	Targeting TLR2.0 with Nme2Cas9	GGGCACCACCCCG GTGAACAGCTC
12	pAE122	Nme2TLR17	pLKO	Targeting TLR2.0 with Nme2Cas9	GGCACCACCCCGGT GAACAGCTCC
13	pAE123	Nme2TLR18	pLKO	Targeting TLR2.0 with Nme2Cas9	GGGATGGGCACCA CCCCGGTGAAC
14	pAE124	Nme2TLR19	pLKO	Targeting TLR2.0 with Nme2Cas9	GCGTGTCCGGCTTT GGCGAGACAA
15	pAE125	Nme2TLR20	pLKO	Targeting TLR2.0 with Nme2Cas9	GTCCGGCTTTGGCG AGACAAATCA
16	pAE126	Nme2TLR21	pLKO	Targeting TLR2.0 with Nme2Cas9	GATCACCTGCCTCG TGGAATACGG
17	pAE149	Nme2TLR22	pLKO	Targeting TLR2.0 with Nme2Cas9	GACGCTGAACTTGT GGCCGTTTAC
18	pAE150	Nme2TLR23	pLKO	Targeting TLR2.0 with Nme2Cas9	GCCAAAGCCGGAC ACGCTGAACTT
19	pAE193	Nme2TLR13 with 23nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GGAACTTGTGGCCG TTTACGTCG
20	pAE194	Nme2TLR13 with 22nt	pLKO	Targeting TLR2.0 with Nme2Cas9	GAACTTGTGGCCGT TTACGTCG

		spacer			
21	pAE195	Nme2TLR13 with 21nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GACTTGTGGCCGTTTACGTCG
22	pAE196	Nme2TLR13 with 20nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GCTTGTGGCCGTTTACGTCG
23	pAE197	Nme2TLR13 with 19nt spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GTTGTGGCCGTTTACGTCG
24	pAE213	Nme2TLR21 with G22 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GTCACCTGCCTCGTGGAATACGG
25	pAE214	Nme2TLR21 with G21 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GCACCTGCCTCGTGGAATACGG
26	pAE215	Nme2TLR21 with G20 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GACCTGCCTCGTGGAATACGG
27	pAE216	Nme2TLR21 with G19 spacer	pLKO	Targeting TLR2.0 with Nme2Cas9	GCCTGCCTCGTGGAATACGG
28	pAE90	Nme2TS1	pLKO	Targeting *AAVS1* with Nme2Cas9	GGTTCTGGGTACTTTTATCTGTCC
29	pAE93	Nme2TS4	pLKO	Targeting *AAVS1* with Nme2Cas9	GTCTGCCTAACAGGAGGTGGGGGT
30	pAE94	Nme2TS5	pLKO	Targeting *AAVS1* with Nme2Cas9	GAATATCAGGAGACTAGGAAGGAG
31	pAE129	Nme2TS6	pLKO	Targeting *LINC01588* with Nme2Cas9	GCCTCCCTGCAGGGCTGCTCCC

32	pAE130	Nme2TS10	pLKO	Targeting *AAVS1* with Nme2Cas9	GAGCTAGTCTTCTTCCTCCAACCC
33	pAE131	Nme2TS11	pLKO	Targeting *AAVS1* with Nme2Cas9	GATCTGTCCCCTCCACCCCACAGT
34	pAE132	Nme2TS12	pLKO	Targeting *AAVS1* with Nme2Cas9	GGCCCAAATGAAAGGAGTGAGAGG
35	pAE133	Nme2TS13	pLKO	Targeting *AAVS1* with Nme2Cas9	GCATCCTCTTGCTTTCTTTGCCTG
36	pAE136	Nme2TS16	pLKO	Targeting *LINC01588* with Nme2Cas9	GGAGTCGCCAGAGGCCGGTGGTGG
37	pAE137	Nme2TS17	pLKO	Targeting *LINC01588* with Nme2Cas9	GCCCAGCGGCCGGATATCAGCTGC
38	pAE138	Nme2TS18	pLKO	Targeting *CYBB* with Nme2Cas9	GGAAGGGAACATATTACTATTGC
39	pAE139	Nme2TS19	pLKO	Targeting *CYBB* with Nme2Cas9	GTGGAGTGGCCTGCTATCAGCTAC
40	pAE140	Nme2TS20	pLKO	Targeting *CYBB* with Nme2Cas9	GAGGAAGGGAACATATTACTATTG
41	pAE141	Nme2TS21	pLKO	Targeting *CYBB* with Nme2Cas9	GTGAATTCTCATCAGCTAAAATGC
42	pAE144	Nme2TS25	pLKO	Targeting *VEGFA* with Nme2Cas9	GCTCACTCACCCACACAGACACAC
43	pAE145	Nme2TS26	pLKO	Targeting *CFTR* with Nme2Cas9	GGAAGAATTTCATTCTGTTCTCAG
44	pAE146	Nme2TS27	pLKO	Targeting *CFTR* with Nme2Cas9	GCTCAGTTTTCCTGGATTATGCCT
45	pAE152	Nme2TS31	pLKO	Targeting *VEGFA* with Nme2Cas9	GCGTTGGAGCGGGGAGAAGGCCAG
46	pAE153	Nme2TS34	pLKO	Targeting *LINC01588* with Nme2Cas9	GGGCCGCGGAGATAGCTGCAGGGC

47	pAE154	Nme2TS35	pLKO	Targeting *LINC01588* with Nme2Cas9	GCCCACCCGGCGGCGCCTCCCTGC
48	pAE155	Nme2TS36	pLKO	Targeting *LINC01588* with Nme2Cas9	GCGTGGCAGCTGATATCCGGCCGC
49	pAE156	Nme2TS37	pLKO	Targeting *LINC01588* with Nme2Cas9	GCCGCGGCGCGACGTGGAGCCAGC
50	pAE157	Nme2TS38	pLKO	Targeting *LINC01588* with Nme2Cas9	GTGCTCCCCAGCCCAAACCGCCGC
51	pAE159	Nme2TS41	pLKO	Targeting *AGA* with Nme2Cas9	GTCAGATTGGCTTGCTCGGAATTG
52	pAE185	Nme2TS44	pLKO	Targeting *VEGFA* with Nme2Cas9	GCTGGGTGAATGGAGCGAGCAGCG
53	pAE186	Nme2TS45	pLKO	Targeting *VEGFA* with Nme2Cas9	GTCCTGGAGTGACCCCTGGCCTTC
54	pAE187	Nme2TS46	pLKO	Targeting *VEGFA* with Nme2Cas9	GATCCTGGAGTGACCCCTGGCCTT
55	pAE188	Nme2TS47	pLKO	Targeting *VEGFA* with Nme2Cas9	GTGTGTCCCTCTCCCCACCCGTCC
56	pAE189	Nme2TS48	pLKO	Targeting *VEGFA* with Nme2Cas9	GTTGGAGCGGGGAGAAGGCCAGGG
57	pAE190	Nme2TS49	pLKO	Targeting *VEGFA* with Nme2Cas9	GCGTTGGAGCGGGGAGAAGGCCAG
58	pAE191	Nme2TS50	pLKO	Targeting *AGA* with Nme2Cas9	GTACCCTCCAATAATTTGGCTGGC
59	pAE192	Nme2TS51	pLKO	Targeting *AGA* with Nme2Cas9	GATAATTTGGCTGGCAATTCCGAG
60	pAE232	TS64_FancJ1	pLKO	Targeting *FANCJ* with Nme2Cas9	GAAAATTGTGATTTCCAGATCCAC
61	pAE233	TS65_FancJ2	pLKO	Targeting *FANCJ* with Nme2Cas9	GAGCAGAAAAAATTGTGATTTCC

62	pAE200	Nme2TS58 (Nme2DS1)	pLKO	Targeting DS in *VEGFA* with Nme2Cas9	GCAGGGGCCAGGTGTCCTTCTCTG
63	pAE201	Nme2TS59 (Nme2DS2)	pLKO	Targeting DS in *VEGFA* with Nme2Cas9	GAATGGCAGGCGGAGGTTGTACTG
64	pAE202	Nme2TS60 (Nme2DS3)	pLKO	Targeting DS in *VEGFA* with Nme2Cas9	GAGTGAGAGAGTGAGAGAGAGACA
65	pAE203	Nme2TS61 (Nme2DS4)	pLKO	Targeting DS in *VEGFA* with Nme2Cas9	GTGAGCAGGCACCTGTGCCAACAT
66	pAE204	Nme2TS62 (Nme2DS5)	pLKO	Targeting DS in *VEGFA* with Nme2Cas9	GCGTGGGGGCTCCGTGCCCCACGC
67	pAE205	Nme2TS63 (Nme2DS6)	pLKO	Targeting DS in *VEGFA* with Nme2Cas9	GCATGGGCAGGGGCTGGGGTGCAC
68	pAE207	SpyDS1	pLKO	Targeting DS in *VEGFA* with SpyCas9	GGGCCAGGTGTCCTTCTCTG
69	pAE208	SpyDS2	pLKO	Targeting DS in *VEGFA* with SpyCas9	GGCAGGCGGAGGTTGTACTG
70	pAE209	SpyDS3	pLKO	Targeting DS in *VEGFA* with SpyCas9	GAGAGAGTGAGAGAGAGACA
71	pAE210	SpyDS4	pLKO	Targeting DS in *VEGFA* with SpyCas9	GCAGGCACCTGTGCCAACAT

72	pAE211	SpyDS5	pLKO	Targeting DS in *VEGFA* with SpyCas9	GGGGGCTCCGTGCCCCACGC
73	pAE212	SpyDS6	pLKO	Targeting DS in *VEGFA* with SpyCas9	GGGCAGGGGCTGGGGTGCAC
74	pAE169	hDeCas9 Wt in AAV backbone	AAV	Nme2Cas9 all-in-one AAV expression with sgRNA cassette	See examples herein.
75	pAE217	hDeCas9 wt in pMSCG7 backbone	pMCSG7	wildtype Nme2Cas9 for bacterial expression	See examples herein.
76	pAE107	2xNLS Nme2Cas9 with HA	pCdest	Nme2Cas9 CMV-driven expression plasmid	See examples herein.
77	pAE127	hDemonCas9 3X NLS in pMSCG7	pMSCG7	Targeting endogenous loci with Nme2Cas9	See examples herein.
78	pAM172	hNme2Cas9 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro	See examples herein.
79	pAM174	nickase hNme2Cas9 D16A 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro	See examples herein.
80	pAM175	nickase hNme2Cas9 H588A 4X	pCVL	Lentivector containing UCOE, SFFV driven	See examples herein.

		NLS with 3XHA		Nme2Cas9 and Puro	
81	pAM177	dead hNme2Cas9 4X NLS with 3XHA	pCVL	Lentivector containing UCOE, SFFV driven Nme2Cas9 and Puro	See examples herein.

Exemplary oligonucleotides			
Number	**Name**	**Sequence**	**Purpose**
1	AAVS1_TIDE1_FW	TGGCTTAGCACCTCTCCAT	TIDE analysis
2	LINC01588_TIDE_FW	AGAGGAGCCTTCTGACTGCTGCAGA	TIDE analysis
3	AAVS1_TIDE2_FW	TCCGTCTTCCTCCACTCC	TIDE analysis
4	NTS55_TIDE_FW	TAGAGAACTGGGTAGTGTG	TIDE analysis
5	VEGF_TIDE3_FW	GTACATGAAGCAACTCCAGTCCCA	TIDE analysis
6	hCFTR_TIDE1_FW	TGGTGATTATGGGAGAACTGGAGC	TIDE analysis
7	AGA_TIDE1_FW	GGCATAAGGAAATCGAAGGTC	TIDE analysis
8	VEGF_TIDE4_FW	ACACGGGCAGCATGGGAATAGTC	TIDE analysis
9	VEGF_TIDE5_FW	CCTGTGTGGCTTTGCTTTGGTCG	TIDE analysis
10	VEGF_TIDE6_FW	GGAGGAAGAGTAGCTCGCCGAGG	TIDE analysis
11	VEGF_TIDE7_FW	AGGGAGAGGGAAGTGTGGGGAAGG	TIDE analysis
12	AAVS1_TIDE1_RV	AGAACTCAGGACCAACTTATTCTG	TIDE analysis
13	LINC01588_TIDE_RV	ATGACAGACACAACCAGAGGGCA	TIDE analysis

14	AAVS1_TIDE2_RV	TAGGAAGGAGGAGGCCTAAG	TIDE analysis
15	NTS55_TIDE_RV	CCAATATTGCATGGGATGG	TIDE analysis
16	VEGF_TIDE3_RV	ATCAAATTCCAGCACCGAGCGC	TIDE analysis
17	hCFTR_TIDE1_RV	ACCATTGAGGACGTTTGTCTCAC	TIDE analysis
18	AGA_TIDE1_RV	CATGTCCTCAAGTCAAGAACAAG	TIDE analysis
19	VEGF_TIDE4_RV	GCTAGGGGAGAGTCCCACTGTCCA	TIDE analysis
20	VEGF_TIDE5_RV	GTAGGGTGTGATGGGAGGCTAAGC	TIDE analysis
21	VEGF_TIDE6_RV	AGACCGAGTGGCAGTGACAGCAAG	TIDE analysis
22	VEGF_TIDE7_RV	GTCTTCCTGCTCTGTGCGCACGAC	TIDE analysis
23	RandomPAM_FW	TAGCGGCCGCTCATGCGCGGCGCATTACCTTTACNNNNNNNNNNGGAT CCTCTAGAGTCG	Protospacer with randomized PAM
24	RandomPAM_RV	ACAGGAAACAGCTATGACCATGAAAGCTTGCATGCCTGCAGGTCGACTCTA GAGGATC	Protospacer with randomized PAM
25	DS2_ON_FW1	ctacacgacgctcttccgatctCCTGGAGCGTGTACGTTGG	Targeted Deep Seq
26	SpyDS2_OT1_FW1	ctacacgacgctcttccgatctCCTGTGGTCCCAGCTACTTG	Targeted Deep Seq
27	SpyDS2_OT2_FW1	ctacacgacgctcttccgatctATCTGCGATGTCCTCGAGG	Targeted Deep Seq
28	SpyDS2_OT3_FW1	ctacacgacgctcttccgatctTGGTGTGCG	Targeted Deep Seq

		CCTCTAACG	
29	SpyDS2_OT4_FW1	ctacacgacgctcttccgatctGGAGTCTTG CTTTGTCACTCAGA	Targeted Deep Seq
30	SpyDS2_OT5_FW1	ctacacgacgctcttccgatctAGCCTAGAC CCAGTCCCAT	Targeted Deep Seq
31	SpyDS2_OT6_FW1	ctacacgacgctcttccgatctGCTGGGCAT AGTAGTGGACT	Targeted Deep Seq
32	SpyDS2_OT7_FW1	ctacacgacgctcttccgatctTGGGGAGGC TGAGACACGA	Targeted Deep Seq
33	SpyDS2_OT8_FW1	ctacacgacgctcttccgatctCTTGGGAGG CTGAGGCAAG	Targeted Deep Seq
34	DS2_ON_RV1	agacgtgtgctcttccgatctCAGGAGGAT GAGAGCCAGG	Targeted Deep Seq
35	SpyDS2_OT1_RV1	agacgtgtgctcttccgatctCAGGGTCTCA CTCTATCACCCA	Targeted Deep Seq
36	SpyDS2_OT2_RV1	agacgtgtgctcttccgatctACTGAATGGG TTGAACTTGGC	Targeted Deep Seq
37	SpyDS2_OT3_RV1	agacgtgtgctcttccgatctGAGACAGAA TCTTGCTCTGTCTCC	Targeted Deep Seq
38	SpyDS2_OT4_RV1	agacgtgtgctcttccgatctTCCCAGCTAC TTGGGAGGC	Targeted Deep Seq
39	SpyDS2_OT5_RV1	agacgtgtgctcttccgatctCCTGCCCAAA TAGGGAAGCAG	Targeted Deep Seq
40	SpyDS2_OT6_RV1	agacgtgtgctcttccgatctTGGCGCCTTA GTCTCTGCTAC	Targeted Deep Seq
41	SpyDS2_OT7_RV1	agacgtgtgctcttccgatctGCATGAGACA CAGTTTCACTCTG	Targeted Deep Seq
42	SpyDS2_OT8_RV1	agacgtgtgctcttccgatctGAGAGAGTCT CACTGCGTTGC	Targeted Deep Seq
43	DS4_ON_FW3	ctacacgacgctcttccgatctTCTCTCACC	Targeted Deep Seq

		CACTGGGCAC	
44	DS4_ON_RV3	agacgtgtgctcttccgatctGCTTCCAGAC GAGTGCAGA	Targeted Deep Seq
45	SpyDS4_OT1_FW1	ctacacgacgctcttccgatctAAGTTTTCA AACCAGAAGAACTACGAC	Targeted Deep Seq
46	SpyDS4_OT2_FW1	ctacacgacgctcttccgatctCCGGTATAA GTCCTGGAGCG	Targeted Deep Seq
47	SpyDS4_OT3_FW1	ctacacgacgctcttccgatctGCCAGGGAG CAATGGCAG	Targeted Deep Seq
48	SpyDS4_OT6_FW1	ctacacgacgctcttccgatctCCTCGAATT CCACGGGGTT	Targeted Deep Seq
49	DS16_ON_FW1	ctacacgacgctcttccgatctGTTGGTGGG AGGGAAGTGAG	Targeted Deep Seq
50	SpyDS6_OT1_FW1	ctacacgacgctcttccgatctGATGGCGGT TGTAGCGGC	Targeted Deep Seq
51	SpyDS6_OT2_FW1	ctacacgacgctcttccgatctCACATAAAC CTATGTTTCAGCAGA	Targeted Deep Seq
52	SpyDS6_OT3_FW1	ctacacgacgctcttccgatctGCTAGTTGG ATTGAAGCAGGGT	Targeted Deep Seq
53	SpyDS6_OT4_FW1	ctacacgacgctcttccgatctTTGAGTGCG GCAGCTTCC	Targeted Deep Seq
54	SpyDS6_OT6_FW1	ctacacgacgctcttccgatctATAACCCTC CCAGGCAAAGTC	Targeted Deep Seq
55	SpyDS6_OT7_FW1	ctacacgacgctcttccgatctAGCCTGCAC ATCTGAGCTC	Targeted Deep Seq
56	SpyDS6_OT8_FW1	ctacacgacgctcttccgatctGGAGCATTG AAGTGCCTGG	Targeted Deep Seq
57	DeDS6_ON_RV1	agacgtgtgctcttccgatctCAGCCTGGGA CCACTGA	Targeted Deep Seq
58	SpyDS6_OT1_RV1	agacgtgtgctcttccgatctCATCCTCGAC	Targeted Deep Seq

		AGTCGCGG	
59	SpyDS6_OT2_RV1	agacgtgtgctcttccgatctGACTGATCAAGTAGAATACTCATGGG	Targeted Deep Seq
60	SpyDS6_OT3_RV1	agacgtgtgctcttccgatctCCCTGCCAGCACTGAAGC	Targeted Deep Seq
61	SpyDS6_OT4_Rv1	agacgtgtgctcttccgatctGGTTCCTATCTTTCTAGACCAGGAGT	Targeted Deep Seq
62	SpyDS6_OT6_RV1	agacgtgtgctcttccgatctAGTGTGGAGGGCTCAGGG	Targeted Deep Seq
63	SpyDS6_OT7_RV1	agacgtgtgctcttccgatctGATGGGCAGAGGAAGGCAA	Targeted Deep Seq
64	SpyDS6_OT8_RV1	agacgtgtgctcttccgatctTCACTCTCATGAGCGTCCCA	Targeted Deep Seq
65	Nme2DS2_OT1_FW1	ctacacgacgctcttccgatctAAGGTTCCTTGCGGTTCGC	Targeted Deep Seq
66	Nme2DS2_OT1_RV1	agacgtgtgctcttccgatctCGCTGCCATTGCTCCCT	Targeted Deep Seq
67	Nme2DS6_OT1_FW1	ctacacgacgctcttccgatctTCTCGCACATTCTTCACGTCC	Targeted Deep Seq
68	Nme2DS6_OT1_RV1	agacgtgtgctcttccgatctAGGAACCTTCCCGACTTAGGG	Targeted Deep Seq
69	Rosa26_ON_FW1	ctacacgacgctcttccgatctCCCGCCCATCTTCTAGAAAGAC	Targeted Deep Seq
70	Rosa26_OT1_FW1	ctacacgacgctcttccgatctTGCCAGGTGAGGGACTGG	Targeted Deep Seq
71	Rosa26_ON_RV1	agacgtgtgctcttccgatctTCTGGGAGTTCTCTGCTGCC	Targeted Deep Seq
72	Rosa26_OT1_RV1	agacgtgtgctcttccgatctTGCCCAACCTTAGCAAGGAG	Targeted Deep Seq
73	pCSK9_ON_FW2	ctacacgacgctcttccgatcttaccttggagcaacg	Targeted Deep Seq

		gcg	
74	PCSK9_ON_RV2	agacgtgtgctcttccgatctcccaggacgaggatg gag	Targeted Deep Seq
75	Tyr_500_FW3	GATAGTCACTCCAGGGGTTG	TIDE analysis
76	Tyr_500_RV3	GTGGTGAACCAATCAGTCCT	TIDE analysis

RNP Delivery for mammalian genome editing

For RNP experiments, the Neon electroporation system was used exactly as described (Amrani et al., 2018). Briefly, 40 picomoles of 3xNLS-Nme2Cas9 along with 50 picomoles of T7-transcribed sgRNA was assembled in buffer R and electroporated using 10 µL Neon tips. After electroporation, cells were plated in pre-warmed 24-well plates containing the appropriate culture media without antibiotics. Electroporation parameters (voltage, width, number of pulses) were 1150 V, 20 ms, 2 pulses for HEK293T cells; 1000 V, 50 ms, 1 pulse for K562 cells.

In vivo AAV8.Nme2Cas9+sgRNA delivery and liver tissue processing

For the AAV8 vector injections, 8-week-old female C57BL/6NJ mice were injected with 4 x10^{11} genome copies per mouse via tail vein, with the sgRNA targeting a validated site in either *Pcsk9* or *Rosa26*. Mice were sacrificed 28 days after vector administration and liver tissues were collected for analysis. Liver tissues were fixed in 4% formalin overnight, embedded in paraffin, sectioned and stained with hematoxylin and eosin (H&E). Blood was drawn from the facial vein at 0, 14 and 28 days post injection, and serum was isolated using a serum separator (BD, Cat. No. 365967) and stored at -80°C until assay. Serum cholesterol level was measured using the Infinity™ colorimetric endpoint assay (Thermo-Scientific) following the manufacturer's protocol and as previously described (Ibraheim et al., 2018). For the anti-PCSK9 Western blot, 40 µg of protein from tissue or 2 ng of Recombinant Mouse PCSK9 Protein (R&D Systems, 9258-SE-020) were loaded onto a MiniPROTEAN® TGX™ Precast Gel (Bio-Rad). The separated bands were transferred onto a PVDF membrane and blocked with 5% Blocking-Grade Blocker solution (Bio-Rad) for 2 hours at room temperature. Next, the membrane was incubated with rabbit anti-GAPDH (Abcam ab9485, 1:2,000) or goat anti-PCSK9 (R&D Systems AF3985, 1:400) antibodies overnight. Membranes were washed in TBST and incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit (Bio-Rad 1706515, 1:4,000), and donkey anti-goat (R&D Systems HAF109, 1:2,000) secondary antibodies for 2 hours at room

temperature. The membranes were washed again in TBST and visualized using Clarity™ western ECL substrate (Bio-Rad) using an M35A XOMAT Processor (Kodak).

Ex vivo AAV6.Nme2Cas9 delivery in mouse zygotes

Zygotes were incubated in 15 µl drops of KSOM (Potassium-Supplemented Simplex Optimized Medium, Millipore, Cat. No. MR-106-D) containing $3x10^9$ or $3x10^8$ GCs of AAV6.Nme2Cas9.sg*Tyr* vector for 5-6 h (4 zygotes in each drop). After incubation, zygotes were rinsed in M2 and transferred to fresh KSOM for overnight culture. The next day, the embryos that advanced to 2-cell stage were transferred into the oviduct of pseudopregnant recipients and allowed to develop to term.

Experimental

Example I

Discovery of Cas9 orthologs with differentially diverged PIDs

Nme1Cas9 peptide sequence was used as a query in BLAST searches to find all Cas9 orthologs in *Neisseria meningitidis* species. Orthologs with >80% identity to Nme1Cas9 were selected for the remainder of this study. The PIDs were then aligned with that of Nme1Cas9 (residues 820-1082) using ClustalW2 and those with clusters of mutations in the PID were selected for further analysis. An unrooted phylogenetic tree of NmeCas9 orthologs was constructed using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Example II

Cloning, expression and purification of Cas9 and Acr orthologs

Examples of plasmids and oligonucleotides used in this study are listed in Table 3. The PIDs of Nme2Cas9 and Nme3Cas9 were ordered as gBlocks (IDT) to replace the PID of Nme1Cas9 using Gibson Assembly (NEB) in the bacterial expression plasmid pMSCG7 (Zhang et al., 2015), which encodes Nme1Cas9 with a 6xHis tag. The construct was transformed into *E. coli*, expressed and purified as previously described (Pawluk et al., 2016). Briefly, Rosetta (DE3) cells containing the respective Cas9 plasmids were grown at 37°C to an OD_{600} of 0.6 and protein expression was induced by 1mM IPTG for 16 hr at 18°C. Cells were harvested and lysed by sonication in lysis buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 5 mM imidazole, 1 mM DTT] supplemented with 1 mg/mL Lysozyme and protease inhibitor cocktail (Sigma). The lysate was then run through a Ni^{2+}-NTA agarose column (Qiagen), and the bound protein was eluted

with 300 mM imidazole and dialyzed into storage buffer [20 mM HEPES-NaOH (pH 7.5), 250 mM NaCl, 1 mM DTT]. For Acr proteins, 6xHis-tagged proteins were expressed in *E. coli* strain BL21 Rosetta (DE3). Cells were grown at 37 °C to an optical density (OD_{600}) of 0.6 in a shaking incubator. The bacterial cultures were cooled to 18°C, and protein expression was induced by adding 1 mM IPTG for overnight expression. The next day, cells were harvested and resuspended in lysis buffer supplemented with 1 mg/mL Lysozyme and protease inhibitor cocktail (Sigma) and protein was purified using the same protocol as for Cas9. The 6xHis tag was removed by incubation of the resin-bound protein with Tobacco Etch Virus (TEV) protease overnight at 4°C to isolate untagged Acrs.

Example III

In vitro PAM discovery assay

A dsDNA target library with randomized PAM sequences was generated by overlapping PCR, with the forward primer containing the 10-nt randomized PAM region. The library was gel-purified and subjected to in vitro cleavage reaction by purified Cas9 along with T7-transcribed sgRNAs. 300 nM Cas9:sgRNA complex was used to cleave 300 nM of the target fragment in 1X NEBuffer 3.1 (NEB) at 37˚C for 1 hr. The reaction was then treated with proteinase K at 50˚C for 10 minutes and run on a 4% agarose/1xTAE gel. The cleavage product was excised, eluted, and cloned using a previously described protocol (Zhang et al., 2012), with modifications. Briefly, DNA ends were repaired, non-templated 2'-deoxyadenosine tails were added, and Y-shaped adapters were ligated. After PCR, the product was quantitated with KAPA Library Quantification Kit and sequenced using a NextSeq 500 (Illumina) to obtain 75 nt paired-end reads. Sequences were analyzed with custom scripts and R.

Example IV

Transfections and mammalian genome editing

Human codon-optimized Nme2Cas9 was cloned by Gibson Assembly into the pCDest2 plasmid backbone previously used for Nme1Cas9 and SpyCas9 expression (Pawluk et al., 2016; Amrani et al., 2018). Transfection of HEK293T and HEK293T-TLR2.0 cells was performed as previously described (Amrani et al., 2018). For Hepa1-6 transfections, Lipofectamine LTX was used to transfect 500ng of all-in-one AAV.sgRNA.Nme2Cas9 plasmid in 24-well plates (~10^5 cells/well), using cells that had been cultured 24 hours before transfection. For K562 cells stably expressing Nme2Cas9 delivered via lentivector (see below), 50,000 – 150,000 cells were

electroporated with 500 ng sgRNA plasmid using 10 µL Neon tips. To measure indels in all cells 72 hr after transfections, cells were harvested and genomic DNA was extracted using the DNaesy Blood and Tissue kit (Qiagen). The targeted locus was amplified by PCR, Sanger-sequenced (Genewiz), and analyzed by TIDE (Brinkman et al., 2014) using the Desktop Genetics web-based interface (http://tide.deskgen.com).

Example V

Lentiviral transduction of K562 cells to stably express Nme2Cas9

K562 cells stably expressing Nme2Cas9 were generated as previously described for Nme1Cas9 (Amrani et al., 2018). For lentivirus production, the lentiviral vector was co-transfected into HEK293T cells along with the packaging plasmids (Addgene 12260 & 12259) in 6-well plates using TransIT-LT1 transfection reagent (Mirus Bio). After 24 hours, culture media was aspirated from the transfected cells and replaced with 1 mL of fresh DMEM. The next day, the supernatant containing the virus was collected and filtered through a 0.45 µm filter. 10 uL of the undiluted supernatant along with 2.5 ug of Polybrene was used to transduce ~10^6 K562 cells in 6-well plates. The transduced cells were selected using media supplemented with 2.5 µg/mL puromycin.

Example VI

RNP Delivery for mammalian genome editing

For RNP experiments, the Neon electroporation system was used exactly as described (Amrani et al., 2018). Briefly, 40 picomoles of 3xNLS-Nme2Cas9 along with 50 picomoles of T7-transcribed sgRNA was assembled in buffer R and electroporated using 10 µL Neon tips. After electroporation, cells were plated in pre-warmed 24-well plates containing the appropriate culture media without antibiotics. Electroporation parameters (voltage, width, number of pulses) were 1150 V, 20 ms, 2 pulses for HEK293T cells; 1000 V, 50 ms, 1 pulse for K562 cells.

Example VII

GUIDE-seq

GUIDE-seq experiments were performed as described previously (Tsai et al., 2014), with minor modifications (Bolukbasi et al., 2015a). Briefly, HEK293T cells were transfected with 200 ng of Cas9 plasmid, 200 ng of sgRNA plasmid, and 7.5 pmol of annealed GUIDE-seq oligonucleotides using Polyfect (Qiagen). Alternatively, Hepa1-6 cells were transfected as described above. Genomic DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen) 72

h after transfection according to the manufacturer's protocol. Library preparation and sequencing were performed exactly as described previously (Bolukbasi et al., 2015a). For analysis, all sequences with up to ten mismatches with the target site, as well as a C in the fifth PAM position (N_4CN), were considered potential off-target sites. Data were analyzed using the Bioconductor package GUIDEseq version 1.1.17 (Zhu et al., 2017).

Example VIII

Targeted deep sequencing and analysis

We used targeted deep sequencing to confirm the results of GUIDE-seq and to measure indel rates with maximal accuracy. We used two-step PCR amplification to produce DNA fragments for each on- and off-target site. For SpyCas9 editing at DS2 and DS6, we selected the top off-target sites based on GUIDE-seq read counts. For SpyCas9 editing at DS4, fewer candidate off-target sites were identified by GUIDE-seq, and only those with NGG (DS4|OT1, DS4|OT3, DS4|OT6) or NGC (DS4|OT2) PAMs were examined by sequencing. In the first step, we used locus-specific primers bearing universal overhangs with ends complementary to the adapters. In the first step, 2x PCR master mix (NEB) was used to generate fragments bearing the overhangs. In the second step, the purified PCR products were amplified with a universal forward primer and indexed reverse primers. Full-size products (~250 bp) were gel-purified and sequenced on an Illumina MiSeq in paired-end mode. MiSeq data analysis was performed as previously described (Pinello et al., 2016; Ibraheim et al., 2018).

Example IX

Off-target analysis using CRISPRseek

Global off-target predictions for TS25 and TS47 were performed using the Bioconductor package CRISPRseek. Minor changes were made to accommodate characteristics of Nme2Cas9 not shared with SpyCas9. Specifically, we used the following changes to: gRNA.size = 24, PAM = "NNNNCC", PAM.size = 6, RNA.PAM.pattern = "NNNNCN", and candidate off-target sites with fewer than 6 mismatches were collected. The top potential off-target sites based on the numbers and positions of mismatches were selected. Genomic DNA from cells targeted by each respective sgRNA was used to amplify each candidate off-target locus and then analyzed by TIDE.

Example X

Mouse strains and embryo collection

All animal experiments were conducted under the guidance of the Institutional Animal Care and Use Committee (IACUC) of the University of Massachusetts Medical School. C57BL/6NJ (Stock No. 005304). Mice were obtained from The Jackson Laboratory. All animals were maintained in a 12 h light cycle. The middle of the light cycle of the day when a mating plug was observed was considered embryonic day 0.5 (E0.5) of gestation. Zygotes were collected at E0.5 by tearing the ampulla with forceps and incubation in M2 medium containing hyaluronidase to remove cumulus cells.

Example XI

In vivo AAV8.Nme2Cas9+sgRNA delivery and liver tissue processing

For the AAV8 vector injections, 8-week-old female C57BL/6NJ mice were injected with 4 x10^{11} genome copies per mouse via tail vein, with the sgRNA targeting a validated site in either *Pcsk9* or *Rosa26*. Mice were sacrificed 28 days after vector administration and liver tissues were collected for analysis. Liver tissues were fixed in 4% formalin overnight, embedded in paraffin, sectioned and stained with hematoxylin and eosin (H&E). Blood was drawn from the facial vein at 0, 14 and 28 days post injection, and serum was isolated using a serum separator (BD, Cat. No. 365967) and stored at -80°C until assay. Serum cholesterol level was measured using the Infinity™ colorimetric endpoint assay (Thermo-Scientific) following the manufacturer's protocol and as previously described (Ibraheim et al., 2018). For the anti-PCSK9 Western blot, 40 μg of protein from tissue or 2 ng of Recombinant Mouse PCSK9 Protein (R&D Systems, 9258-SE-020) were loaded onto a MiniPROTEAN® TGX™ Precast Gel (Bio-Rad). The separated bands were transferred onto a PVDF membrane and blocked with 5% Blocking-Grade Blocker solution (Bio-Rad) for 2 hours at room temperature. Next, the membrane was incubated with rabbit anti-GAPDH (Abcam ab9485, 1:2,000) or goat anti-PCSK9 (R&D Systems AF3985, 1:400) antibodies overnight. Membranes were washed in TBST and incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit (Bio-Rad 1706515, 1:4,000), and donkey anti-goat (R&D Systems HAF109, 1:2,000) secondary antibodies for 2 hours at room temperature. The membranes were washed again in TBST and visualized using Clarity™ western ECL substrate (Bio-Rad) using an M35A XOMAT Processor (Kodak).

Example XII

Ex vivo AAV6.Nme2Cas9 delivery in mouse zygotes

Zygotes were incubated in 15 μl drops of KSOM (Potassium-Supplemented Simplex

Optimized Medium, Millipore, Cat. No. MR-106-D) containing $3x10^9$ or $3x10^8$ GCs of AAV6.Nme2Cas9.sg*Tyr* vector for 5-6 h (4 zygotes in each drop). After incubation, zygotes were rinsed in M2 and transferred to fresh KSOM for overnight culture. The next day, the embryos that advanced to 2-cell stage were transferred into the oviduct of pseudopregnant recipients and allowed to develop to term.

REFERENCES, each of which are herein incorporated by reference in their entirety:

Amrani, N., Gao, X.D., Liu, P., Edraki, A., Mir, A., Ibraheim, R., Gupta, A., Sasaki, K.E., Wu, T., Donohoue, P.D., et al. (2018). NmeCas9 is an intrinsically high-fidelity genome editing platform. BioRxiv, https://doi.org/10.1101/172650.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Bisaria, N., Jarmoskaite, I., and Herschlag, D. (2017). Lessons from Enzyme Kinetics Reveal Specificity Principles for RNA-Guided Nucleases in RNA Interference and CRISPR-Based Genome Editing. Cell Syst. 4, 21-29.

Bolukbasi, M.F., Gupta, A., Oikemus, S., Derr, A.G., Garber, M., Brodsky, M.H., Zhu, L.J., and Wolfe, S.A. (2015a). DNA-binding-domain fusions enhance the targeting range and precision of Cas9. Nat. Methods 12, 1150-1156.

Bolukbasi, M.F., Gupta, A., and Wolfe, S.A. (2015b). Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. Nat. Methods 13, 41-50.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 42, e168.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960-964.

Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., et al. (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nat. Biotechnol. 36, 265-271.

Certo, M.T., Ryu, B.Y., Annis, J.E., Garibov, M., Jarjour, J., Rawlings, D.J., and Scharenberg, A.M. (2011). Tracking genome engineering outcome at individual DNA breakpoints. Nat. Methods 8, 671-676.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550, 407-410.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 31, 230-232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 132-141.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. 190, 1390-1400.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. 17, 5-15.

Dong, Guo, M., Wang, S., Zhu, Y., Wang, S., Xiong, Z., Yang, J., Xu, Z., and Huang, Z. (2017). Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. Nature 546, 436-439.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116-1121.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lecrivain, A.L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 42, 2577-2590.

Friedland, A.E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., Marco, E., Gotta, G.M., Maeder, M.L., Kennedy, E.M., et al. (2015). Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome Biol. 16, 257.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev. 5, 1513-1523.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 32, 279-284.

Gallagher, D.N., and Haber, J.E. (2018). Repair of a Site-Specific DNA Cleavage: Old-School Lessons for Cas9-Mediated Gene Editing. ACS Chem. Biol. 13, 397-405.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA 109, E2579-2586.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature 551, 464-471.

Ghanta, K., Dokshin, G., Mir, A., Krishnamurthy, P., Gneid, H., Edraki, A., Watts, J., Sontheimer, E., and Mello, C. (2018). 5′ Modifications Improve Potency and Efficacy of DNA Donors for Precision Genome Editing. Biorxiv 354480.

Gorski, S.A., Vogel, J., and Doudna, J.A. (2017). RNA-based recognition and targeting: sowing the seeds of specificity. Nat. Rev. Mol. Cell Biol. 18, 215-228.

Harrington, L.B., Doxzen, K.W., Ma, E., Liu, J.J., Knott, G.J., Edraki, A., Garcia, B., Amrani, N., Chen, J.S., Cofsky, J.C., et al. (2017a). A Broad-Spectrum Inhibitor of CRISPR-Cas9. Cell 170, 1224-1233.

Harrington, L.B., Paez-Espino, D., Staahl, B.T., Chen, J.S., Ma, E., Kyrpides, N.C., and Doudna, J.A. (2017b). A thermostable Cas9 with increased lifetime in human plasma. Nat. Commun. 8, 1424.

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. USA 110, 15644-15649.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57-63.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227-229.

Hynes, A.P., Rousseau, G.M., Lemay, M.-L., Horvath, P., Romero, D.A., Fremaux, C., and Moineau, S. (2017). An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. Nat. Microbiol. 2, 1374-1380.

Ibraheim, R., Song, C.-Q., Mir, A., Amrani, N., Xue, W., and Sontheimer, E.J. (2018). All-in-One Adeno-associated Virus Delivery and Genome Editing by Neisseria meningitidis Cas9 in vivo. BioRxiv, https://doi.org/10.1101/295055.

Jiang, F., and Doudna, J.A. (2017). CRISPR–Cas9 Structures and Mechanisms. Annu. Rev. Biophys. 46, 505-529.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. eLife 2, e00471.

Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253.

Keeler, A.M., ElMallah, M.K., and Flotte, T.R. (2017). Gene Therapy 2017: Progress and Future Directions. Clin. Transl. Sci. 10, 242-248.

Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K.-E., Kim, K., Cho, H.-Y., Song, D.W., Lee, K.J., Jung, M.H., et al. (2017). In vivo genome editing with a small Cas9 ortholog derived from Campylobacter jejuni. Nat. Commun. 8, 14500.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 24, 1012-1019.

Kim, B., Komor, A., Levy, J., Packer, M., Zhao, K., and Liu, D. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nature Biotechnology 35.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. 33, 1293-1298.

Kluesner, M., Nedveck, D., Lahr, W., Garbe, J., Abrahante, J., Webber, B., and Moriarity, B. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. The CRISPR Journal 1, 239–250.

Koblan, L., Doman, J., Wilson, C., Levy, J., Tay, T., Newby, G., Maianti, J., Raguram, A., and Liu, D. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat Biotechnol 36, 843.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20-36.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424.

Lee, C.M., Cradick, T.J., and Bao, G. (2016). The Neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells. Mol. Ther. 24, 645-654.

Lee, J., Mir, A., Edraki, A., Garcia, B., Amrani, N., Lou, H.E., Gainetdinov, I., Pawluk, A., Ibraheim, R., Gao, X.D., et al. (2018). Potent Cas9 inhibition in bacterial and human cells by new anti-CRISPR protein families. BioRxiv, https://www.biorxiv.org/content/early/2018/2006/2020/350504.

Ma, E., Harrington, L.B., O'Connell, M.R., Zhou, K., and Doudna, J.A. (2015). Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. Mol. Cell 60, 398-407.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833-838.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-guided human genome engineering via Cas9. Science 339, 823-826.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843-1845.

Mir, A., Edraki, A., Lee, J., and Sontheimer, E.J. (2018). Type II-C CRISPR-Cas9 biology, mechanism and application. ACS Chem. Biol. 13, 357-365.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733-740.

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., and Banfield, J.F. (2015). CRISPR immunity drives rapid phage genome evolution in Streptococcus thermophilus. mBio 6.

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L., et al. (2016). Naturally occurring off-switches for CRISPR-Cas9. Cell 167, 1829-1838 e1829.

Pawluk, A., Bondy-Denomy, J., Cheung, V.H., Maxwell, K.L., and Davidson, A.R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. mBio 5, e00896.

Pinello, L., Canver, M.C., Hoban, M.D., Orkin, S.H., Kohn, D.B., Bauer, D.E., and Yuan, G.C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. Nat. Biotechnol. 34, 695-697.

Racanelli, V., and Rehermann, B. (2006). The liver as an immunological organ. Hepatology 43, S54-62.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-191.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380-1389.

Rashid, S., Curtis, D.E., Garuti, R., Anderson, N.N., Bashmakov, Y., Ho, Y.K., Hammer, R.E., Moon, Y.A., and Horton, J.D. (2005). Decreased plasma cholesterol and hypersensitivity to statins in mice lacking Pcsk9. Proc. Natl. Acad. Sci. USA 102, 5374-5379.

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell 168, 150-158 e110.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 39, 9275-9282.

Schumann, K., Lin, S., Boyer, E., Simeonov, D.R., Subramaniam, M., Gate, R.E., Haliburton, G.E., Ye, C.J., Bluestone, J.A., Doudna, J.A., et al. (2015). Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. Proc. Natl. Acad. Sci. USA 112, 10437-10442.

Shin, J., Jiang, F., Liu, J.J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Sci. Adv. 3, e1701620.

Tsai, S.Q., and Joung, J.K. (2016). Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. Nat. Rev. Genet. 17, 300-312.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2014). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187-197.

Tycko, J., Myer, V.E., and Hsu, P.D. (2016). Methods for optimizing CRISPR-Cas9 genome editing specificity. Mol. Cell 63, 355-370.

Yang, H., and Patel, D.J. (2017). Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9. Mol Cell 67, 117-127 e115.

Yin, H., Song, C.Q., Suresh, S., Kwan, S.Y., Wu, Q., Walsh, S., Ding, J., Bogorad, R.L., Zhu, L.J., Wolfe, S.A., et al. (2018). Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. Nat. Chem. Biol. 14, 311-316.

Yokoyama, T., Silversides, D.W., Waymire, K.G., Kwon, B.S., Takeuchi, T., and Overbeek, P.A. (1990). Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice. Nucleic Acids Res. 18, 7293-7298.

Yoon, Y., Wang, D., Tai, P.W.L., Riley, J., Gao, G., and Rivera-Perez, J.A. (2018). Streamlined ex vivo and in vivo genome editing in mouse embryos using recombinant adeno-associated viruses. Nat. Commun. 9, 412.

Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., Vogel, J., and Sontheimer, E.J. (2013). Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. Mol. Cell 50, 488-503.

Zhang, Y., Rajan, R., Seifert, H.S., Mondragón, A., and Sontheimer, E.J. (2015). DNase H activity of Neisseria meningitidis Cas9. Mol. Cell 60, 242-255.

Zhang, Z., Theurkauf, W.E., Weng, Z., and Zamore, P.D. (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence 3, 9.

Zhu, L.J., Holmes, B.R., Aronin, N., and Brodsky, M.H. (2014). CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS One 9, e108424.

Zhu, L.J., Lawrence, M., Gupta, A., Pagés, H., Kucukural, A., Garber, M., and Wolfe, S.A. (2017). GUIDEseq: a bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases. BMC Genomics 18, 379.

Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.-Y., and Liu, D.R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat. Biotechnol. 33, 73-80.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in biological control, biochemistry, molecular

biology, entomology, plankton, fishery systems, and fresh water ecology, or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

<110> University of Massachusetts

<120> Programmable DNA Base Editing by Nme2cas9-Deaminase Fusion Proteins

<130> P000900SG2

<150> 11202103722T
<151> 2019-10-15

<150> PCT/US2019/056341
<151> 2019-10-15

<150> US 62/745,666
<151> 2018-10-15

<160> 971

<170> PatentIn version 3.5

<210> 1
<211> 1425
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 1

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 Tyr 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 Glu Asn Arg
115 120 125

```
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
    130                 135                 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145                 150                 155                 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
                165                 170                 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
            180                 185                 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
        195                 200                 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
    210                 215                 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225                 230                 235                 240

Ala Thr Pro Glu Ser Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr
                245                 250                 255

Ile Leu Gly Leu Ala Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val
            260                 265                 270

Glu Ile Asp Glu Glu Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val
        275                 280                 285

Arg Val Phe Glu Arg Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala
    290                 295                 300

Met Ala Arg Arg Leu Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg
305                 310                 315                 320

Ala His Arg Leu Leu Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val
                325                 330                 335

Leu Gln Ala Ala Asp Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro
            340                 345                 350

Asn Thr Pro Trp Gln Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr
        355                 360                 365
```

```
Pro Leu Glu Trp Ser Ala Val Leu Leu His Leu Ile Lys His Arg Gly
    370                 375                 380

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

Gly Ala Leu Leu Lys Gly Val Ala Asn Asn Ala His Ala Leu Gln Thr
                405                 410                 415

Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys
            420                 425                 430

Glu Ser Gly His Ile Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe
        435                 440                 445

Ser Arg Lys Asp Leu Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln
    450                 455                 460

Lys Glu Phe Gly Asn Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile
465                 470                 475                 480

Glu Thr Leu Leu Met Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val
                485                 490                 495

Gln Lys Met Leu Gly His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala
            500                 505                 510

Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu
        515                 520                 525

Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp
    530                 535                 540

Thr Glu Arg Ala Thr Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu
545                 550                 555                 560

Thr Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe
                565                 570                 575

Phe Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu
            580                 585                 590

Met Glu Met Lys Ala Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu
        595                 600                 605
```

```
Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln
    610                 615                 620

Asp Glu Ile Gly Thr Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile
625                 630                 635                 640

Thr Gly Arg Leu Lys Asp Arg Val Gln Pro Glu Ile Leu Glu Ala Leu
                645                 650                 655

Leu Lys His Ile Ser Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala
            660                 665                 670

Leu Arg Arg Ile Val Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu
        675                 680                 685

Ala Cys Ala Glu Ile Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu
    690                 695                 700

Glu Lys Ile Tyr Leu Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro
705                 710                 715                 720

Val Val Leu Arg Ala Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val
                725                 730                 735

Val Arg Arg Tyr Gly Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg
            740                 745                 750

Glu Val Gly Lys Ser Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln
        755                 760                 765

Glu Glu Asn Arg Lys Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu
    770                 775                 780

Tyr Phe Pro Asn Phe Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys
785                 790                 795                 800

Leu Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys
                805                 810                 815

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

Ala Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys
        835                 840                 845

Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro
```

```
    850                 855                 860

Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe
865                 870                 875                 880

Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg
                885                 890                 895

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

Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp
        915                 920                 925

His Ile Leu Leu Thr Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn
    930                 935                 940

Gly Gln Ile Thr Asn Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val
945                 950                 955                 960

Arg Ala Glu Asn Asp Arg His His Ala Leu Asp Ala Val Val Val Ala
                965                 970                 975

Cys Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr
            980                 985                 990

Lys Glu Met Asn Ala Phe Asp Gly  Lys Thr Ile Asp Lys  Glu Thr Gly
        995                 1000                1005

Lys Val  Leu His Gln Lys Thr  His Phe Pro Gln Pro  Trp Glu Phe
    1010                1015                1020

Phe Ala  Gln Glu Val Met Ile  Arg Val Phe Gly Lys  Pro Asp Gly
    1025                1030                1035

Lys Pro  Glu Phe Glu Glu Ala  Asp Thr Pro Glu Lys  Leu Arg Thr
    1040                1045                1050

Leu Leu  Ala Glu Lys Leu Ser  Ser Arg Pro Glu Ala  Val His Glu
    1055                1060                1065

Tyr Val  Thr Pro Leu Phe Val  Ser Arg Ala Pro Asn  Arg Lys Met
    1070                1075                1080

Ser Gly  Ala His Lys Asp Thr  Leu Arg Ser Ala Lys  Arg Phe Val
    1085                1090                1095
```

```
Lys His  Asn Glu Lys Ile Ser  Val Lys Arg Val Trp  Leu Thr Glu
    1100                1105                1110

Ile Lys  Leu Ala Asp Leu Glu  Asn Met Val Asn Tyr  Lys Asn Gly
    1115                1120                1125

Arg Glu  Ile Glu Leu Tyr Glu  Ala Leu Lys Ala Arg  Leu Glu Ala
    1130                1135                1140

Tyr Gly  Gly Asn Ala Lys Gln  Ala Phe Asp Pro Lys  Asp Asn Pro
    1145                1150                1155

Phe Tyr  Lys Lys Gly Gly Gln  Leu Val Lys Ala Val  Arg Val Glu
    1160                1165                1170

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

Thr Ile  Ala Asp Asn Gly Asp  Met Val Arg Val Asp  Val Phe Cys
    1190                1195                1200

Lys Val  Asp Lys Lys Gly Lys  Asn Gln Tyr Phe Ile  Val Pro Ile
    1205                1210                1215

Tyr Ala  Trp Gln Val Ala Glu  Asn Ile Leu Pro Asp  Ile Asp Cys
    1220                1225                1230

Lys Gly  Tyr Arg Ile Asp Asp  Ser Tyr Thr Phe Cys  Phe Ser Leu
    1235                1240                1245

His Lys  Tyr Asp Leu Ile Ala  Phe Gln Lys Asp Glu  Lys Ser Lys
    1250                1255                1260

Val Glu  Phe Ala Tyr Tyr Ile  Asn Cys Asp Ser Ser  Asn Gly Arg
    1265                1270                1275

Phe Tyr  Leu Ala Trp His Asp  Lys Gly Ser Lys Glu  Gln Gln Phe
    1280                1285                1290

Arg Ile  Ser Thr Gln Asn Leu  Val Leu Ile Gln Lys  Tyr Gln Val
    1295                1300                1305

Asn Glu  Leu Gly Lys Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg
    1310                1315                1320
```

```
Pro Pro  Val Arg Ser Gly Gly  Ser Thr Asn Leu Ser  Asp Ile Ile
    1325                1330                1335

Glu Lys  Glu Thr Gly Lys Gln  Leu Val Ile Gln Glu  Ser Ile Leu
    1340                1345                1350

Met Leu  Pro Glu Glu Val Glu  Glu Val Ile Gly Asn  Lys Pro Glu
    1355                1360                1365

Ser Asp  Ile Leu Val His Thr  Ala Tyr Asp Glu Ser  Thr Asp Glu
    1370                1375                1380

Asn Val  Met Leu Leu Thr Ser  Asp Ala Pro Glu Tyr  Lys Pro Trp
    1385                1390                1395

Ala Leu  Val Ile Gln Asp Ser  Asn Gly Glu Asn Lys  Ile Lys Met
    1400                1405                1410

Leu Ser  Gly Gly Ser Pro Lys  Lys Lys Arg Lys Val
    1415                1420                1425
```

<210> 2
<211> 1425
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 2

```
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 Tyr 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 Glu Asn Arg
        115                 120                 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
    130                 135                 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145                 150                 155                 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
                165                 170                 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
            180                 185                 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
        195                 200                 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
    210                 215                 220

Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225                 230                 235                 240

Ala Thr Pro Glu Ser Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr
                245                 250                 255

Ile Leu Gly Leu Ala Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val
            260                 265                 270

Glu Ile Asp Glu Glu Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val
        275                 280                 285

Arg Val Phe Glu Arg Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala
    290                 295                 300

Met Ala Arg Arg Leu Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg
305                 310                 315                 320

Ala His Arg Leu Leu Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val
                325                 330                 335
```

```
Leu Gln Ala Ala Asp Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro
            340                 345                 350

Asn Thr Pro Trp Gln Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr
        355                 360                 365

Pro Leu Glu Trp Ser Ala Val Leu Leu His Leu Ile Lys His Arg Gly
    370                 375                 380

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

Gly Ala Leu Leu Lys Gly Val Ala Asn Asn Ala His Ala Leu Gln Thr
                405                 410                 415

Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys
            420                 425                 430

Glu Ser Gly His Ile Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe
        435                 440                 445

Ser Arg Lys Asp Leu Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln
    450                 455                 460

Lys Glu Phe Gly Asn Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile
465                 470                 475                 480

Glu Thr Leu Leu Met Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val
                485                 490                 495

Gln Lys Met Leu Gly His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala
            500                 505                 510

Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu
        515                 520                 525

Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp
    530                 535                 540

Thr Glu Arg Ala Thr Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu
545                 550                 555                 560

Thr Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe
                565                 570                 575
```

Phe Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu
580 585 590

Met Glu Met Lys Ala Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu
595 600 605

Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln
610 615 620

Asp Glu Ile Gly Thr Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile
625 630 635 640

Thr Gly Arg Leu Lys Asp Arg Val Gln Pro Glu Ile Leu Glu Ala Leu
645 650 655

Leu Lys His Ile Ser Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala
660 665 670

Leu Arg Arg Ile Val Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu
675 680 685

Ala Cys Ala Glu Ile Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu
690 695 700

Glu Lys Ile Tyr Leu Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro
705 710 715 720

Val Val Leu Arg Ala Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val
725 730 735

Val Arg Arg Tyr Gly Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg
740 745 750

Glu Val Gly Lys Ser Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln
755 760 765

Glu Glu Asn Arg Lys Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu
770 775 780

Tyr Phe Pro Asn Phe Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys
785 790 795 800

Leu Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys
805 810 815

Glu Ile Asn Leu Val Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp

820 825 830

Ala Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys
835 840 845

Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro
850 855 860

Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe
865 870 875 880

Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg
885 890 895

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

Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp
915 920 925

His Ile Leu Leu Thr Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn
930 935 940

Gly Gln Ile Thr Asn Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val
945 950 955 960

Arg Ala Glu Asn Asp Arg His His Ala Leu Asp Ala Val Val Val Ala
965 970 975

Cys Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr
980 985 990

Lys Glu Met Asn Ala Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly
995 1000 1005

Lys Val Leu His Gln Lys Thr His Phe Pro Gln Pro Trp Glu Phe
1010 1015 1020

Phe Ala Gln Glu Val Met Ile Arg Val Phe Gly Lys Pro Asp Gly
1025 1030 1035

Lys Pro Glu Phe Glu Glu Ala Asp Thr Pro Glu Lys Leu Arg Thr
1040 1045 1050

Leu Leu Ala Glu Lys Leu Ser Ser Arg Pro Glu Ala Val His Glu
1055 1060 1065

```
Tyr Val  Thr Pro Leu Phe Val  Ser Arg Ala Pro Asn  Arg Lys Met
    1070                1075                1080

Ser Gly  Ala His Lys Asp Thr  Leu Arg Ser Ala Lys  Arg Phe Val
    1085                1090                1095

Lys His  Asn Glu Lys Ile Ser  Val Lys Arg Val Trp  Leu Thr Glu
    1100                1105                1110

Ile Lys  Leu Ala Asp Leu Glu  Asn Met Val Asn Tyr  Lys Asn Gly
    1115                1120                1125

Arg Glu  Ile Glu Leu Tyr Glu  Ala Leu Lys Ala Arg  Leu Glu Ala
    1130                1135                1140

Tyr Gly  Gly Asn Ala Lys Gln  Ala Phe Asp Pro Lys  Asp Asn Pro
    1145                1150                1155

Phe Tyr  Lys Lys Gly Gly Gln  Leu Val Lys Ala Val  Arg Val Glu
    1160                1165                1170

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

Thr Ile  Ala Asp Asn Gly Asp  Met Val Arg Val Asp  Val Phe Cys
    1190                1195                1200

Lys Val  Asp Lys Lys Gly Lys  Asn Gln Tyr Phe Ile  Val Pro Ile
    1205                1210                1215

Tyr Ala  Trp Gln Val Ala Glu  Asn Ile Leu Pro Asp  Ile Asp Cys
    1220                1225                1230

Lys Gly  Tyr Arg Ile Asp Asp  Ser Tyr Thr Phe Cys  Phe Ser Leu
    1235                1240                1245

His Lys  Tyr Asp Leu Ile Ala  Phe Gln Lys Asp Glu  Lys Ser Lys
    1250                1255                1260

Val Glu  Phe Ala Tyr Tyr Ile  Asn Cys Asp Ser Ser  Asn Gly Arg
    1265                1270                1275

Phe Tyr  Leu Ala Trp His Asp  Lys Gly Ser Lys Glu  Gln Gln Phe
    1280                1285                1290
```

```
Arg Ile  Ser Thr Gln Asn Leu  Val Leu Ile Gln Lys  Tyr Gln Val
    1295                1300                1305

Asn Glu  Leu Gly Lys Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg
    1310                1315                1320

Pro Pro  Val Arg Ser Gly Gly  Ser Thr Asn Leu Ser  Asp Ile Ile
    1325                1330                1335

Glu Lys  Glu Thr Gly Lys Gln  Leu Val Ile Gln Glu  Ser Ile Leu
    1340                1345                1350

Met Leu  Pro Glu Glu Val Glu  Glu Val Ile Gly Asn  Lys Pro Glu
    1355                1360                1365

Ser Asp  Ile Leu Val His Thr  Ala Tyr Asp Glu Ser  Thr Asp Glu
    1370                1375                1380

Asn Val  Met Leu Leu Thr Ser  Asp Ala Pro Glu Tyr  Lys Pro Trp
    1385                1390                1395

Ala Leu  Val Ile Gln Asp Ser  Asn Gly Glu Asn Lys  Ile Lys Met
    1400                1405                1410

Leu Ser  Gly Gly Ser Pro Lys  Lys Lys Arg Lys Val
    1415                1420                1425
```

<210> 3
<211> 1497
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 3

```
Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1               5                  10                  15

Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
            20                  25                  30

Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
        35                  40                  45

Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
    50                  55                  60
```

```
Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65                  70                  75                  80

Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95

Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
            100                 105                 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
        115                 120                 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    130                 135                 140

Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145                 150                 155                 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190

Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
    290                 295                 300
```

```
Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Met Ala Ala
385                 390                 395                 400

Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile
                405                 410                 415

Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu Glu Asn Pro
            420                 425                 430

Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg Ala Glu Val
        435                 440                 445

Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu Ala Arg Ser
    450                 455                 460

Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala Arg
465                 470                 475                 480

Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp Phe Asp Glu
                485                 490                 495

Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala
            500                 505                 510

Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu
        515                 520                 525

Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu
    530                 535                 540
```

```
Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala
545                 550                 555                 560

Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu
                565                 570                 575

Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln
            580                 585                 590

Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu
        595                 600                 605

Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val
    610                 615                 620

Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg
625                 630                 635                 640

Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly His Cys Thr
                645                 650                 655

Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr Ala Glu
            660                 665                 670

Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Gln
        675                 680                 685

Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu Met Asp
    690                 695                 700

Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg Lys Leu
705                 710                 715                 720

Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr Gly Lys
                725                 730                 735

Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala Tyr His Ala
            740                 745                 750

Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro
        755                 760                 765

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

Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val
```

785 790 795 800

Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser Phe Asp Lys
805 810 815

Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val Pro Leu Met
820 825 830

Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp
835 840 845

His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile
850 855 860

Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser Gln
865 870 875 880

Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro Ala
885 890 895

Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp
900 905 910

Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu
915 920 925

Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu
930 935 940

Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His
945 950 955 960

Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg Leu Asn
965 970 975

Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr
980 985 990

Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn
995 1000 1005

Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys
1010 1015 1020

Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu Thr
1025 1030 1035

```
Ser Arg  Phe Pro Arg Ser Lys  Lys Gln Arg Ile Leu  Leu Gln Lys
    1040                1045                1050

Phe Asp  Glu Asp Gly Phe Lys  Glu Cys Asn Leu Asn  Asp Thr Arg
    1055                1060                1065

Tyr Val  Asn Arg Phe Leu Cys  Gln Phe Val Ala Asp  His Ile Leu
    1070                1075                1080

Leu Thr  Gly Lys Gly Lys Arg  Arg Val Phe Ala Ser  Asn Gly Gln
    1085                1090                1095

Ile Thr  Asn Leu Leu Arg Gly  Phe Trp Gly Leu Arg  Lys Val Arg
    1100                1105                1110

Ala Glu  Asn Asp Arg His His  Ala Leu Asp Ala Val  Val Val Ala
    1115                1120                1125

Cys Ser  Thr Val Ala Met Gln  Gln Lys Ile Thr Arg  Phe Val Arg
    1130                1135                1140

Tyr Lys  Glu Met Asn Ala Phe  Asp Gly Lys Thr Ile  Asp Lys Glu
    1145                1150                1155

Thr Gly  Lys Val Leu His Gln  Lys Thr His Phe Pro  Gln Pro Trp
    1160                1165                1170

Glu Phe  Phe Ala Gln Glu Val  Met Ile Arg Val Phe  Gly Lys Pro
    1175                1180                1185

Asp Gly  Lys Pro Glu Phe Glu  Glu Ala Asp Thr Pro  Glu Lys Leu
    1190                1195                1200

Arg Thr  Leu Leu Ala Glu Lys  Leu Ser Ser Arg Pro  Glu Ala Val
    1205                1210                1215

His Glu  Tyr Val Thr Pro Leu  Phe Val Ser Arg Ala  Pro Asn Arg
    1220                1225                1230

Lys Met  Ser Gly Ala His Lys  Asp Thr Leu Arg Ser  Ala Lys Arg
    1235                1240                1245

Phe Val  Lys His Asn Glu Lys  Ile Ser Val Lys Arg  Val Trp Leu
    1250                1255                1260
```

```
Thr Glu  Ile Lys Leu Ala Asp  Leu Glu Asn Met Val  Asn Tyr Lys
    1265                1270                1275

Asn Gly  Arg Glu Ile Glu Leu  Tyr Glu Ala Leu Lys  Ala Arg Leu
    1280                1285                1290

Glu Ala  Tyr Gly Gly Asn Ala  Lys Gln Ala Phe Asp  Pro Lys Asp
    1295                1300                1305

Asn Pro  Phe Tyr Lys Lys Gly  Gly Gln Leu Val Lys  Ala Val Arg
    1310                1315                1320

Val Glu  Lys Thr Gln Glu Ser  Gly Val Leu Leu Asn  Lys Lys Asn
    1325                1330                1335

Ala Tyr  Thr Ile Ala Asp Asn  Gly Asp Met Val Arg  Val Asp Val
    1340                1345                1350

Phe Cys  Lys Val Asp Lys Lys  Gly Lys Asn Gln Tyr  Phe Ile Val
    1355                1360                1365

Pro Ile  Tyr Ala Trp Gln Val  Ala Glu Asn Ile Leu  Pro Asp Ile
    1370                1375                1380

Asp Cys  Lys Gly Tyr Arg Ile  Asp Asp Ser Tyr Thr  Phe Cys Phe
    1385                1390                1395

Ser Leu  His Lys Tyr Asp Leu  Ile Ala Phe Gln Lys  Asp Glu Lys
    1400                1405                1410

Ser Lys  Val Glu Phe Ala Tyr  Tyr Ile Asn Cys Asp  Ser Ser Asn
    1415                1420                1425

Gly Arg  Phe Tyr Leu Ala Trp  His Asp Lys Gly Ser  Lys Glu Gln
    1430                1435                1440

Gln Phe  Arg Ile Ser Thr Gln  Asn Leu Val Leu Ile  Gln Lys Tyr
    1445                1450                1455

Gln Val  Asn Glu Leu Gly Lys  Glu Ile Arg Pro Cys  Arg Leu Lys
    1460                1465                1470

Lys Arg  Pro Pro Val Arg Glu  Asp Lys Arg Pro Ala  Ala Thr Lys
    1475                1480                1485
```

```
Lys Ala  Gly Gln Ala Lys Lys  Lys Lys
    1490                1495


<210>  4
<211>  1497
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Synthetic

<400>  4

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1               5                  10                  15


Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val Gly Ala
            20                  25                  30


Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Pro
        35                  40                  45


Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
    50                  55                  60


Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65                  70                  75                  80


Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
                85                  90                  95


Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys Thr Gly
            100                 105                 110


Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His
        115                 120                 125


Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
    130                 135                 140


Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala Gln Lys
145                 150                 155                 160


Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser
                165                 170                 175


Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            180                 185                 190
```

```
Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His Glu Tyr
        195                 200                 205

Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg
    210                 215                 220

Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly
225                 230                 235                 240

Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala
                245                 250                 255

Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg
            260                 265                 270

Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys
        275                 280                 285

Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val
    290                 295                 300

Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His
305                 310                 315                 320

Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala
                325                 330                 335

Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln
            340                 345                 350

Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly
        355                 360                 365

Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
    370                 375                 380

Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Met Ala Ala
385                 390                 395                 400

Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile
                405                 410                 415

Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu Glu Asn Pro
            420                 425                 430
```

```
Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg Ala Glu Val
        435                 440                 445

Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu Ala Arg Ser
    450                 455                 460

Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala Arg
465                 470                 475                 480

Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp Phe Asp Glu
                485                 490                 495

Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala
            500                 505                 510

Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu
        515                 520                 525

Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu
    530                 535                 540

Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala
545                 550                 555                 560

Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu
                565                 570                 575

Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln
            580                 585                 590

Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu
        595                 600                 605

Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val
    610                 615                 620

Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg
625                 630                 635                 640

Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly His Cys Thr
                645                 650                 655

Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr Ala Glu
            660                 665                 670

Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Gln
```

675 680 685

Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu Met Asp
690 695 700

Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg Lys Leu
705 710 715 720

Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr Gly Lys
725 730 735

Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala Tyr His Ala
740 745 750

Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro
755 760 765

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

Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val
785 790 795 800

Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser Phe Asp Lys
805 810 815

Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val Pro Leu Met
820 825 830

Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp
835 840 845

His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile
850 855 860

Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser Gln
865 870 875 880

Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro Ala
885 890 895

Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp
900 905 910

Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu
915 920 925

```
Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu
    930                 935                 940

Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His
945                 950                 955                 960

Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg Leu Asn
                965                 970                 975

Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr
            980                 985                 990

Trp Asp Asp Ser Phe Asn Asn Lys  Val Leu Val Leu Gly  Ser Glu Asn
        995                 1000                1005

Gln Asn  Lys Gly Asn Gln Thr  Pro Tyr Glu Tyr Phe  Asn Gly Lys
    1010                1015                1020

Asp Asn  Ser Arg Glu Trp Gln  Glu Phe Lys Ala Arg  Val Glu Thr
    1025                1030                1035

Ser Arg  Phe Pro Arg Ser Lys  Lys Gln Arg Ile Leu  Leu Gln Lys
    1040                1045                1050

Phe Asp  Glu Asp Gly Phe Lys  Glu Cys Asn Leu Asn  Asp Thr Arg
    1055                1060                1065

Tyr Val  Asn Arg Phe Leu Cys  Gln Phe Val Ala Asp  His Ile Leu
    1070                1075                1080

Leu Thr  Gly Lys Gly Lys Arg  Arg Val Phe Ala Ser  Asn Gly Gln
    1085                1090                1095

Ile Thr  Asn Leu Leu Arg Gly  Phe Trp Gly Leu Arg  Lys Val Arg
    1100                1105                1110

Ala Glu  Asn Asp Arg His His  Ala Leu Asp Ala Val  Val Val Ala
    1115                1120                1125

Cys Ser  Thr Val Ala Met Gln  Gln Lys Ile Thr Arg  Phe Val Arg
    1130                1135                1140

Tyr Lys  Glu Met Asn Ala Phe  Asp Gly Lys Thr Ile  Asp Lys Glu
    1145                1150                1155
```

```
Thr Gly  Lys Val Leu His Gln  Lys Thr His Phe Pro  Gln Pro Trp
    1160                1165                1170

Glu Phe  Phe Ala Gln Glu Val  Met Ile Arg Val Phe  Gly Lys Pro
    1175                1180                1185

Asp Gly  Lys Pro Glu Phe Glu  Glu Ala Asp Thr Pro  Glu Lys Leu
    1190                1195                1200

Arg Thr  Leu Leu Ala Glu Lys  Leu Ser Ser Arg Pro  Glu Ala Val
    1205                1210                1215

His Glu  Tyr Val Thr Pro Leu  Phe Val Ser Arg Ala  Pro Asn Arg
    1220                1225                1230

Lys Met  Ser Gly Ala His Lys  Asp Thr Leu Arg Ser  Ala Lys Arg
    1235                1240                1245

Phe Val  Lys His Asn Glu Lys  Ile Ser Val Lys Arg  Val Trp Leu
    1250                1255                1260

Thr Glu  Ile Lys Leu Ala Asp  Leu Glu Asn Met Val  Asn Tyr Lys
    1265                1270                1275

Asn Gly  Arg Glu Ile Glu Leu  Tyr Glu Ala Leu Lys  Ala Arg Leu
    1280                1285                1290

Glu Ala  Tyr Gly Gly Asn Ala  Lys Gln Ala Phe Asp  Pro Lys Asp
    1295                1300                1305

Asn Pro  Phe Tyr Lys Lys Gly  Gly Gln Leu Val Lys  Ala Val Arg
    1310                1315                1320

Val Glu  Lys Thr Gln Glu Ser  Gly Val Leu Leu Asn  Lys Lys Asn
    1325                1330                1335

Ala Tyr  Thr Ile Ala Asp Asn  Gly Asp Met Val Arg  Val Asp Val
    1340                1345                1350

Phe Cys  Lys Val Asp Lys Lys  Gly Lys Asn Gln Tyr  Phe Ile Val
    1355                1360                1365

Pro Ile  Tyr Ala Trp Gln Val  Ala Glu Asn Ile Leu  Pro Asp Ile
    1370                1375                1380
```

```
Asp Cys  Lys Gly Tyr Arg Ile  Asp Asp Ser Tyr Thr  Phe Cys Phe
    1385                1390                1395

Ser Leu  His Lys Tyr Asp Leu  Ile Ala Phe Gln Lys  Asp Glu Lys
    1400                1405                1410

Ser Lys  Val Glu Phe Ala Tyr  Tyr Ile Asn Cys Asp  Ser Ser Asn
    1415                1420                1425

Gly Arg  Phe Tyr Leu Ala Trp  His Asp Lys Gly Ser  Lys Glu Gln
    1430                1435                1440

Gln Phe  Arg Ile Ser Thr Gln  Asn Leu Val Leu Ile  Gln Lys Tyr
    1445                1450                1455

Gln Val  Asn Glu Leu Gly Lys  Glu Ile Arg Pro Cys  Arg Leu Lys
    1460                1465                1470

Lys Arg  Pro Pro Val Arg Glu  Asp Lys Arg Pro Ala  Ala Thr Lys
    1475                1480                1485

Lys Ala  Gly Gln Ala Lys Lys  Lys Lys
    1490                1495
```

<210> 5
<211> 1524
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 5

```
Met Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1               5                  10                  15

Arg Lys Val Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His
            20                  25                  30

Ala Leu Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro Val
        35                  40                  45

Gly Ala Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp Asn
    50                  55                  60

Arg Pro Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met Ala
65                  70                  75                  80
```

```
Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala
                85                  90                  95

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

Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala Lys
        115                 120                 125

Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met
    130                 135                 140

Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala
145                 150                 155                 160

Ala Leu Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys Ala
                165                 170                 175

Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly
            180                 185                 190

Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
        195                 200                 205

Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser His
    210                 215                 220

Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp
225                 230                 235                 240

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg Val
                245                 250                 255

Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala
            260                 265                 270

His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn
        275                 280                 285

Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val
    290                 295                 300

Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe
305                 310                 315                 320
```

```
Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val
                325                 330                 335

Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile
            340                 345                 350

Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met Pro
        355                 360                 365

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

Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser
385                 390                 395                 400

Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Met
                405                 410                 415

Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Ala Ile
            420                 425                 430

Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu Glu
        435                 440                 445

Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg Ala
    450                 455                 460

Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu Ala
465                 470                 475                 480

Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg
                485                 490                 495

Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp Phe
            500                 505                 510

Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu
        515                 520                 525

Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala
    530                 535                 540

Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg Lys
545                 550                 555                 560

Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly
```

565 570 575

Val Ala Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro
580 585 590

Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg
595 600 605

Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln
610 615 620

Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro
625 630 635 640

His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met Thr
645 650 655

Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly His
660 665 670

Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr
675 680 685

Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu
690 695 700

Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu
705 710 715 720

Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg
725 730 735

Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr
740 745 750

Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala Tyr
755 760 765

His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys
770 775 780

Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr Ala
785 790 795 800

Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp
805 810 815

```
Arg Val Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser Phe
            820                 825                 830

Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val Pro
        835                 840                 845

Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr
    850                 855                 860

Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro
865                 870                 875                 880

Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu
                885                 890                 895

Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser
            900                 905                 910

Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe
        915                 920                 925

Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp
    930                 935                 940

Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val
945                 950                 955                 960

Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln
                965                 970                 975

Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg
            980                 985                 990

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

Arg Thr  Trp Asp Asp Ser Phe  Asn Asn Lys Val Leu  Val Leu Gly
    1010                1015                1020

Ser Glu  Asn Gln Asn Lys Gly  Asn Gln Thr Pro Tyr  Glu Tyr Phe
    1025                1030                1035

Asn Gly  Lys Asp Asn Ser Arg  Glu Trp Gln Glu Phe  Lys Ala Arg
    1040                1045                1050
```

```
Val Glu  Thr Ser Arg Phe Pro  Arg Ser Lys Lys Gln  Arg Ile Leu
    1055                1060                1065

Leu Gln  Lys Phe Asp Glu Asp  Gly Phe Lys Glu Cys  Asn Leu Asn
    1070                1075                1080

Asp Thr  Arg Tyr Val Asn Arg  Phe Leu Cys Gln Phe  Val Ala Asp
    1085                1090                1095

His Ile  Leu Leu Thr Gly Lys  Gly Lys Arg Arg Val  Phe Ala Ser
    1100                1105                1110

Asn Gly  Gln Ile Thr Asn Leu  Leu Arg Gly Phe Trp  Gly Leu Arg
    1115                1120                1125

Lys Val  Arg Ala Glu Asn Asp  Arg His His Ala Leu  Asp Ala Val
    1130                1135                1140

Val Val  Ala Cys Ser Thr Val  Ala Met Gln Gln Lys  Ile Thr Arg
    1145                1150                1155

Phe Val  Arg Tyr Lys Glu Met  Asn Ala Phe Asp Gly  Lys Thr Ile
    1160                1165                1170

Asp Lys  Glu Thr Gly Lys Val  Leu His Gln Lys Thr  His Phe Pro
    1175                1180                1185

Gln Pro  Trp Glu Phe Phe Ala  Gln Glu Val Met Ile  Arg Val Phe
    1190                1195                1200

Gly Lys  Pro Asp Gly Lys Pro  Glu Phe Glu Glu Ala  Asp Thr Pro
    1205                1210                1215

Glu Lys  Leu Arg Thr Leu Leu  Ala Glu Lys Leu Ser  Ser Arg Pro
    1220                1225                1230

Glu Ala  Val His Glu Tyr Val  Thr Pro Leu Phe Val  Ser Arg Ala
    1235                1240                1245

Pro Asn  Arg Lys Met Ser Gly  Ala His Lys Asp Thr  Leu Arg Ser
    1250                1255                1260

Ala Lys  Arg Phe Val Lys His  Asn Glu Lys Ile Ser  Val Lys Arg
    1265                1270                1275
```

```
Val Trp  Leu Thr Glu Ile Lys  Leu Ala Asp Leu Glu  Asn Met Val
    1280                 1285                 1290

Asn Tyr  Lys Asn Gly Arg Glu  Ile Glu Leu Tyr Glu  Ala Leu Lys
    1295                 1300                 1305

Ala Arg  Leu Glu Ala Tyr Gly  Gly Asn Ala Lys Gln  Ala Phe Asp
    1310                 1315                 1320

Pro Lys  Asp Asn Pro Phe Tyr  Lys Lys Gly Gly Gln  Leu Val Lys
    1325                 1330                 1335

Ala Val  Arg Val Glu Lys Thr  Gln Glu Ser Gly Val  Leu Leu Asn
    1340                 1345                 1350

Lys Lys  Asn Ala Tyr Thr Ile  Ala Asp Asn Gly Asp  Met Val Arg
    1355                 1360                 1365

Val Asp  Val Phe Cys Lys Val  Asp Lys Lys Gly Lys  Asn Gln Tyr
    1370                 1375                 1380

Phe Ile  Val Pro Ile Tyr Ala  Trp Gln Val Ala Glu  Asn Ile Leu
    1385                 1390                 1395

Pro Asp  Ile Asp Cys Lys Gly  Tyr Arg Ile Asp Asp  Ser Tyr Thr
    1400                 1405                 1410

Phe Cys  Phe Ser Leu His Lys  Tyr Asp Leu Ile Ala  Phe Gln Lys
    1415                 1420                 1425

Asp Glu  Lys Ser Lys Val Glu  Phe Ala Tyr Tyr Ile  Asn Cys Asp
    1430                 1435                 1440

Ser Ser  Asn Gly Arg Phe Tyr  Leu Ala Trp His Asp  Lys Gly Ser
    1445                 1450                 1455

Lys Glu  Gln Gln Phe Arg Ile  Ser Thr Gln Asn Leu  Val Leu Ile
    1460                 1465                 1470

Gln Lys  Tyr Gln Val Asn Glu  Leu Gly Lys Glu Ile  Arg Pro Cys
    1475                 1480                 1485

Arg Leu  Lys Lys Arg Pro Pro  Val Arg Glu Asp Lys  Arg Pro Ala
    1490                 1495                 1500

Ala Thr  Lys Lys Ala Gly Gln  Ala Lys Lys Lys Lys  Phe Glu Pro
```

1505 1510 1515

Lys Lys Lys Arg Lys Val
1520

<210> 6
<211> 1697
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 6

Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Gly Ser Gly Gly Gly Ser
1 5 10 15

Gly Gly Gly Ser Gly Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Gly
20 25 30

Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Pro Leu Glu Pro Lys Lys
35 40 45

Lys Arg Lys Val Pro Trp Ser Ser Glu Thr Gly Pro Val Ala Val Asp
50 55 60

Pro Thr Leu Arg Arg Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe
65 70 75 80

Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn
85 90 95

Trp Gly Gly Arg His Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn
100 105 110

Lys His Val Glu Val Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr
115 120 125

Phe Cys Pro Asn Thr Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser
130 135 140

Pro Cys Gly Glu Cys Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr
145 150 155 160

Pro His Val Thr Leu Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala
165 170 175

```
Asp Pro Arg Asn Arg Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val
            180                 185                 190

Thr Ile Gln Ile Met Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn
        195                 200                 205

Phe Val Asn Tyr Ser Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro
    210                 215                 220

His Leu Trp Val Arg Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu
225                 230                 235                 240

Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu
                245                 250                 255

Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro
            260                 265                 270

Pro His Ile Leu Trp Ala Thr Gly Leu Lys Ser Gly Gly Ser Ser Gly
        275                 280                 285

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
    290                 295                 300

Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Ile Asp Lys Leu Ala Ala
305                 310                 315                 320

Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile
                325                 330                 335

Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu Glu Asn Pro
            340                 345                 350

Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg Ala Glu Val
        355                 360                 365

Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu Ala Arg Ser
    370                 375                 380

Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala Arg
385                 390                 395                 400

Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp Phe Asp Glu
                405                 410                 415

Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala
```

420 425 430

Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu
435 440 445

Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu
450 455 460

Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala
465 470 475 480

Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu
485 490 495

Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln
500 505 510

Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu
515 520 525

Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val
530 535 540

Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg
545 550 555 560

Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly His Cys Thr
565 570 575

Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr Ala Glu
580 585 590

Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Gln
595 600 605

Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu Met Asp
610 615 620

Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg Lys Leu
625 630 635 640

Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr Gly Lys
645 650 655

Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala Tyr His Ala
660 665 670

```
Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro
        675                 680                 685

Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr Ala Phe Ser
    690                 695                 700

Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val
705                 710                 715                 720

Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser Phe Asp Lys
                725                 730                 735

Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val Pro Leu Met
            740                 745                 750

Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp
        755                 760                 765

His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile
    770                 775                 780

Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser Gln
785                 790                 795                 800

Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro Ala
                805                 810                 815

Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp
            820                 825                 830

Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu
        835                 840                 845

Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu
    850                 855                 860

Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His
865                 870                 875                 880

Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg Leu Asn
                885                 890                 895

Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr
            900                 905                 910
```

```
Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn
        915                 920                 925

Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp
    930                 935                 940

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

Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys Phe Asp Glu
                965                 970                 975

Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr Val Asn Arg
            980                 985                 990

Phe Leu Cys Gln Phe Val Ala Asp  His Ile Leu Leu Thr  Gly Lys Gly
        995                 1000                1005

Lys Arg  Arg Val Phe Ala Ser  Asn Gly Gln Ile Thr  Asn Leu Leu
    1010                1015                1020

Arg Gly  Phe Trp Gly Leu Arg  Lys Val Arg Ala Glu  Asn Asp Arg
    1025                1030                1035

His His  Ala Leu Asp Ala Val  Val Val Ala Cys Ser  Thr Val Ala
    1040                1045                1050

Met Gln  Gln Lys Ile Thr Arg  Phe Val Arg Tyr Lys  Glu Met Asn
    1055                1060                1065

Ala Phe  Asp Gly Lys Thr Ile  Asp Lys Glu Thr Gly  Lys Val Leu
    1070                1075                1080

His Gln  Lys Thr His Phe Pro  Gln Pro Trp Glu Phe  Phe Ala Gln
    1085                1090                1095

Glu Val  Met Ile Arg Val Phe  Gly Lys Pro Asp Gly  Lys Pro Glu
    1100                1105                1110

Phe Glu  Glu Ala Asp Thr Pro  Glu Lys Leu Arg Thr  Leu Leu Ala
    1115                1120                1125

Glu Lys  Leu Ser Ser Arg Pro  Glu Ala Val His Glu  Tyr Val Thr
    1130                1135                1140
```

```
Pro Leu  Phe Val Ser Arg Ala  Pro Asn Arg Lys Met  Ser Gly Ala
    1145                1150                1155

His Lys  Asp Thr Leu Arg Ser  Ala Lys Arg Phe Val  Lys His Asn
    1160                1165                1170

Glu Lys  Ile Ser Val Lys Arg  Val Trp Leu Thr Glu  Ile Lys Leu
    1175                1180                1185

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

Glu Leu  Tyr Glu Ala Leu Lys  Ala Arg Leu Glu Ala  Tyr Gly Gly
    1205                1210                1215

Asn Ala  Lys Gln Ala Phe Asp  Pro Lys Asp Asn Pro  Phe Tyr Lys
    1220                1225                1230

Lys Gly  Gly Gln Leu Val Lys  Ala Val Arg Val Glu  Lys Thr Gln
    1235                1240                1245

Glu Ser  Gly Val Leu Leu Asn  Lys Lys Asn Ala Tyr  Thr Ile Ala
    1250                1255                1260

Asp Asn  Gly Asp Met Val Arg  Val Asp Val Phe Cys  Lys Val Asp
    1265                1270                1275

Lys Lys  Gly Lys Asn Gln Tyr  Phe Ile Val Pro Ile  Tyr Ala Trp
    1280                1285                1290

Gln Val  Ala Glu Asn Ile Leu  Pro Asp Ile Asp Cys  Lys Gly Tyr
    1295                1300                1305

Arg Ile  Asp Asp Ser Tyr Thr  Phe Cys Phe Ser Leu  His Lys Tyr
    1310                1315                1320

Asp Leu  Ile Ala Phe Gln Lys  Asp Glu Lys Ser Lys  Val Glu Phe
    1325                1330                1335

Ala Tyr  Tyr Ile Asn Cys Asp  Ser Ser Asn Gly Arg  Phe Tyr Leu
    1340                1345                1350

Ala Trp  His Asp Lys Gly Ser  Lys Glu Gln Gln Phe  Arg Ile Ser
    1355                1360                1365

Thr Gln  Asn Leu Val Leu Ile  Gln Lys Tyr Gln Val  Asn Glu Leu
```

1370 1375 1380

Gly Lys Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val
1385 1390 1395

Arg Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Tyr Pro Tyr
1400 1405 1410

Asp Val Pro Asp Tyr Ala Gly Ser Tyr Pro Tyr Asp Val Pro Asp
1415 1420 1425

Tyr Ala Gly Ser Ala Ala Pro Ala Ala Lys Lys Lys Lys Leu Asp
1430 1435 1440

Phe Glu Ser Gly Glu Phe Leu Gln Pro Gly Ile Asp Leu Ser Gln
1445 1450 1455

Leu Gly Gly Asp Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr
1460 1465 1470

Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val
1475 1480 1485

Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val
1490 1495 1500

Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr
1505 1510 1515

Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala
1520 1525 1530

Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly
1535 1540 1545

Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Gly Gly Ser Gly
1550 1555 1560

Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys
1565 1570 1575

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val
1580 1585 1590

Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His
1595 1600 1605

```
Thr Ala  Tyr Asp Glu Ser Thr  Asp Glu Asn Val Met  Leu Leu Thr
    1610                1615                1620

Ser Asp  Ala Pro Glu Tyr Lys  Pro Trp Ala Leu Val  Ile Gln Asp
    1625                1630                1635

Ser Asn  Gly Glu Asn Lys Ile  Lys Met Leu Ser Gly  Gly Ser Pro
    1640                1645                1650

Lys Lys  Lys Arg Lys Val Ser  Arg Gly Ser Ala Ala  Pro Ala Ala
    1655                1660                1665

Lys Arg  Val Lys Leu Asp Gly  Gly Ser Gly Gly Gly  Ser Gly Gly
    1670                1675                1680

Gly Ser  Gly Ser Gly Pro Ala  Ala Lys Arg Val Lys  Leu Asp
    1685                1690                1695
```

<210> 7
<211> 1641
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 7

```
Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Gly Ser Gly Gly Gly Ser
1               5                  10                  15

Gly Gly Gly Ser Gly Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Gly
            20                  25                  30

Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Pro Leu Glu Pro Lys Lys
        35                  40                  45

Lys Arg Lys Val Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg
    50                  55                  60

His Ala Leu Thr Leu Ala Lys Arg Ala Trp Asp Glu Arg Glu Val Pro
65                  70                  75                  80

Val Gly Ala Val Leu Val His Asn Asn Arg Val Ile Gly Glu Gly Trp
                85                  90                  95

Asn Arg Pro Ile Gly Arg His Asp Pro Thr Ala His Ala Glu Ile Met
```

```
           100                 105                 110

Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp
        115                 120                 125

Ala Thr Leu Tyr Val Thr Leu Glu Pro Cys Val Met Cys Ala Gly Ala
    130                 135                 140

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Ala Arg Asp Ala
145                 150                 155                 160

Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly
                165                 170                 175

Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys
            180                 185                 190

Ala Ala Leu Leu Ser Asp Phe Phe Arg Met Arg Arg Gln Glu Ile Lys
        195                 200                 205

Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly
    210                 215                 220

Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro
225                 230                 235                 240

Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Glu Val Glu Phe Ser
                245                 250                 255

His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg
            260                 265                 270

Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg
        275                 280                 285

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr
    290                 295                 300

Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln
305                 310                 315                 320

Asn Tyr Arg Leu Ile Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys
                325                 330                 335

Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val
            340                 345                 350
```

```
Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp
        355                 360                 365

Val Leu His Tyr Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly
    370                 375                 380

Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Tyr Phe Phe Arg Met
385                 390                 395                 400

Pro Arg Gln Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp
                405                 410                 415

Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr
            420                 425                 430

Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser Gly Gly Ser
        435                 440                 445

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Asp Ile Asp Lys Leu
    450                 455                 460

Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Ala Ile
465                 470                 475                 480

Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu Glu
                485                 490                 495

Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg Ala
            500                 505                 510

Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu Ala
        515                 520                 525

Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg
    530                 535                 540

Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp Phe
545                 550                 555                 560

Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu
                565                 570                 575

Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala
            580                 585                 590
```

```
Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg Lys
        595                 600                 605

Asn Glu Gly Glu Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly
    610                 615                 620

Val Ala Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro
625                 630                 635                 640

Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg
                645                 650                 655

Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln
            660                 665                 670

Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro
        675                 680                 685

His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met Thr
    690                 695                 700

Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly His
705                 710                 715                 720

Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr
                725                 730                 735

Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu
            740                 745                 750

Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu
        755                 760                 765

Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg
    770                 775                 780

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

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

His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys
            820                 825                 830
```

```
Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr Ala
        835                 840                 845

Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp
    850                 855                 860

Arg Val Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser Phe
865                 870                 875                 880

Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val Pro
                885                 890                 895

Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr
            900                 905                 910

Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro
        915                 920                 925

Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu
    930                 935                 940

Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser
945                 950                 955                 960

Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe
                965                 970                 975

Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp
            980                 985                 990

Arg Glu Lys Ala Ala Ala Lys Phe  Arg Glu Tyr Phe Pro  Asn Phe Val
        995                 1000                1005

Gly Glu  Pro Lys Ser Lys Asp  Ile Leu Lys Leu Arg  Leu Tyr Glu
    1010                1015                1020

Gln Gln  His Gly Lys Cys Leu  Tyr Ser Gly Lys Glu  Ile Asn Leu
    1025                1030                1035

Val Arg  Leu Asn Glu Lys Gly  Tyr Val Glu Ile Asp  His Ala Leu
    1040                1045                1050

Pro Phe  Ser Arg Thr Trp Asp  Asp Ser Phe Asn Asn  Lys Val Leu
    1055                1060                1065

Val Leu  Gly Ser Glu Asn Gln  Asn Lys Gly Asn Gln  Thr Pro Tyr
```

```
    1070                1075                1080

Glu Tyr  Phe Asn Gly Lys Asp  Asn Ser Arg Glu Trp  Gln Glu Phe
    1085                1090                1095

Lys Ala  Arg Val Glu Thr Ser  Arg Phe Pro Arg Ser  Lys Lys Gln
    1100                1105                1110

Arg Ile  Leu Leu Gln Lys Phe  Asp Glu Asp Gly Phe  Lys Glu Cys
    1115                1120                1125

Asn Leu  Asn Asp Thr Arg Tyr  Val Asn Arg Phe Leu  Cys Gln Phe
    1130                1135                1140

Val Ala  Asp His Ile Leu Leu  Thr Gly Lys Gly Lys  Arg Arg Val
    1145                1150                1155

Phe Ala  Ser Asn Gly Gln Ile  Thr Asn Leu Leu Arg  Gly Phe Trp
    1160                1165                1170

Gly Leu  Arg Lys Val Arg Ala  Glu Asn Asp Arg His  His Ala Leu
    1175                1180                1185

Asp Ala  Val Val Val Ala Cys  Ser Thr Val Ala Met  Gln Gln Lys
    1190                1195                1200

Ile Thr  Arg Phe Val Arg Tyr  Lys Glu Met Asn Ala  Phe Asp Gly
    1205                1210                1215

Lys Thr  Ile Asp Lys Glu Thr  Gly Lys Val Leu His  Gln Lys Thr
    1220                1225                1230

His Phe  Pro Gln Pro Trp Glu  Phe Phe Ala Gln Glu  Val Met Ile
    1235                1240                1245

Arg Val  Phe Gly Lys Pro Asp  Gly Lys Pro Glu Phe  Glu Glu Ala
    1250                1255                1260

Asp Thr  Pro Glu Lys Leu Arg  Thr Leu Leu Ala Glu  Lys Leu Ser
    1265                1270                1275

Ser Arg  Pro Glu Ala Val His  Glu Tyr Val Thr Pro  Leu Phe Val
    1280                1285                1290

Ser Arg  Ala Pro Asn Arg Lys  Met Ser Gly Ala His  Lys Asp Thr
    1295                1300                1305
```

```
Leu Arg  Ser Ala Lys Arg Phe  Val Lys His Asn Glu  Lys Ile Ser
    1310                1315                1320

Val Lys  Arg Val Trp Leu Thr  Glu Ile Lys Leu Ala  Asp Leu Glu
    1325                1330                1335

Asn Met  Val Asn Tyr Lys Asn  Gly Arg Glu Ile Glu  Leu Tyr Glu
    1340                1345                1350

Ala Leu  Lys Ala Arg Leu Glu  Ala Tyr Gly Gly Asn  Ala Lys Gln
    1355                1360                1365

Ala Phe  Asp Pro Lys Asp Asn  Pro Phe Tyr Lys Lys  Gly Gly Gln
    1370                1375                1380

Leu Val  Lys Ala Val Arg Val  Glu Lys Thr Gln Glu  Ser Gly Val
    1385                1390                1395

Leu Leu  Asn Lys Lys Asn Ala  Tyr Thr Ile Ala Asp  Asn Gly Asp
    1400                1405                1410

Met Val  Arg Val Asp Val Phe  Cys Lys Val Asp Lys  Lys Gly Lys
    1415                1420                1425

Asn Gln  Tyr Phe Ile Val Pro  Ile Tyr Ala Trp Gln  Val Ala Glu
    1430                1435                1440

Asn Ile  Leu Pro Asp Ile Asp  Cys Lys Gly Tyr Arg  Ile Asp Asp
    1445                1450                1455

Ser Tyr  Thr Phe Cys Phe Ser  Leu His Lys Tyr Asp  Leu Ile Ala
    1460                1465                1470

Phe Gln  Lys Asp Glu Lys Ser  Lys Val Glu Phe Ala  Tyr Tyr Ile
    1475                1480                1485

Asn Cys  Asp Ser Ser Asn Gly  Arg Phe Tyr Leu Ala  Trp His Asp
    1490                1495                1500

Lys Gly  Ser Lys Glu Gln Gln  Phe Arg Ile Ser Thr  Gln Asn Leu
    1505                1510                1515

Val Leu  Ile Gln Lys Tyr Gln  Val Asn Glu Leu Gly  Lys Glu Ile
    1520                1525                1530
```

Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg Val Tyr Pro
1535 1540 1545

Tyr Asp Val Pro Asp Tyr Ala Gly Tyr Pro Tyr Asp Val Pro Asp
1550 1555 1560

Tyr Ala Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Ser
1565 1570 1575

Ala Ala Pro Ala Ala Lys Lys Lys Leu Asp Phe Glu Ser Gly
1580 1585 1590

Glu Phe Leu Gln Pro Gly Gly Ser Thr Ser Ser Arg Gly Ser Ala
1595 1600 1605

Ala Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Gly Ser Gly Gly
1610 1615 1620

Gly Ser Gly Gly Gly Ser Gly Ser Gly Pro Ala Ala Lys Arg Val
1625 1630 1635

Lys Leu Asp
1640

<210> 8
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 8
ggttctgggt actttatct gtcc 24

<210> 9
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 9
cctccacc 8

<210> 10
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 10
tggcttagca cctctccat 19

<210> 11
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 11
agaactcagg accaacttat tctg 24

<210> 12
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 12
gtctgcctaa caggaggtgg gggt 24

<210> 13
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 13
tagacgaa 8

<210> 14
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 14
tggcttagca cctctccat 19

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 15
agaactcagg accaacttat tctg 24

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 16
gaatatcagg agactaggaa ggag 24

<210> 17
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 17
gaggccta 8

<210> 18
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 18
tggcttagca cctctccat 19

<210> 19
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 19
agaactcagg accaacttat tctg 24

<210> 20
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 20
gcctccctgc agggctgctc cc 22

<210> 21
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 21
cagcccaa 8

<210> 22
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 22
aggagccttc tgactgctgc a 21

<210> 23
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 23
atgacagaca caaccagagg gca 23

<210> 24
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 24
gagctagtct tcttcctcca accc 24

<210> 25
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 25
gggcccta 8

<210> 26
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 26
tggcttagca cctctccat 19

<210> 27
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 27
agaactcagg accaacttat tctg 24

<210> 28
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 28
gatctgtccc ctccacccca cagt 24

<210> 29
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 29
ggggccac 8

<210> 30
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 30
tggcttagca cctctccat 19

<210> 31
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 31
agaactcagg accaacttat tctg 24

<210> 32
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 32
ggcccaaatg aaaggagtga gagg 24

<210> 33
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 33
tgacccga 8

<210> 34
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 34
tccgtcttcc tccactcc 18

<210> 35
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 35

taggaaggag gaggcctaag 20

<210> 36
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 36
gcatcctctt gctttctttg cctg 24

<210> 37
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 37
gacacccc 8

<210> 38
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 38
tccgtcttcc tccactcc 18

<210> 39
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 39
taggaaggag gaggcctaag 20

<210> 40
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 40
ggagtcgcca gaggccggtg gtgg 24

<210> 41
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 41
atttcctc 8

<210> 42
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 42
aggagccttc tgactgctgc a 21

<210> 43
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 43
atgacagaca caaccagagg gca 23

<210> 44
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 44
gcccagcggc cggatatcag ctgc 24

<210> 45
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 45
cacgcccg 8

<210> 46
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 46
aggagccttc tgactgctgc a 21

<210> 47
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 47
atgacagaca caaccagagg gca 23

<210> 48
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 48
ggaagggaac atattactat tgc 23

<210> 49
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 49
tttccctc 8

<210> 50
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 50
tagagaactg ggtagtgtg 19

<210> 51
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 51
ccaatattgc atgggatgg 19

<210> 52
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 52
gtggagtggc ctgctatcag ctac 24

<210> 53
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 53
ctatccaa 8

<210> 54
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 54
tagagaactg ggtagtgtg 19

<210> 55
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 55
ccaatattgc atgggatgg 19

<210> 56

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 56
gaggaaggga acatattact attg 24

<210> 57
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 57
ctttccct 8

<210> 58
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 58
tagagaactg ggtagtgtg 19

<210> 59
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 59
ccaatattgc atgggatgg 19

<210> 60
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 60
gtgaattctc atcagctaaa atgc 24

<210> 61
<211> 8

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 61
caagcctt 8

<210> 62
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 62
tagagaactg ggtagtgtg 19

<210> 63
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 63
ccaatattgc atgggatgg 19

<210> 64
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 64
gctcactcac ccacacagac acac 24

<210> 65
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 65
acgtcctc 8

<210> 66
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 66
tacatgaagc aactccagtc cc 22

<210> 67
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 67
atcaaattcc agcaccgagc gc 22

<210> 68
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 68
ggaagaattt cattctgttc tcag 24

<210> 69
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 69
ttttcctg 8

<210> 70
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 70
ggtgattatg ggagaactgg ag 22

<210> 71
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 71
accattgagg acgtttgtct cac 23

<210> 72
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 72
gctcagtttt cctggattat gcct 24

<210> 73
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 73
ggcaccat 8

<210> 74
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 74
ggtgattatg ggagaactgg ag 22

<210> 75
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 75
accattgagg acgtttgtct cac 23

<210> 76
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 76
gcgttggagc ggggagaagg ccag 24

<210> 77
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 77
gggtcact 8

<210> 78
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 78
tacatgaagc aactccagtc cc 22

<210> 79
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 79
atcaaattcc agcaccgagc gc 22

<210> 80
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 80
gggccgcgga gatagctgca gggc 24

<210> 81
<211> 8
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 81
ggggcccc 8

<210> 82
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 82
aggagccttc tgactgctgc a 21

<210> 83
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 83
atgacagaca caaccagagg gca 23

<210> 84
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 84
gcccacccgg cggcgcctcc ctgc 24

<210> 85
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 85
agggctgc 8

<210> 86
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 86
aggagccttc tgactgctgc a 21

<210> 87
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 87
atgacagaca caaccagagg gca 23

<210> 88
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 88
gcgtggcagc tgatatccgg ccgc 24

<210> 89
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 89
tgggcgtc 8

<210> 90
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 90
aggagccttc tgactgctgc a 21

<210> 91
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 91
atgacagaca caaccagagg gca 23

<210> 92
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 92
gccgcggcgc gacgtggagc cagc 24

<210> 93
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 93
cccgcaaa 8

<210> 94
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 94
aggagccttc tgactgctgc a 21

<210> 95
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 95
atgacagaca caaccagagg gca 23

<210> 96
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 96

gtgctcccca gcccaaaccg ccgc 24

<210> 97
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 97
ggcgcgac 8

<210> 98
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 98
aggagccttc tgactgctgc a 21

<210> 99
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 99
atgacagaca caaccagagg gca 23

<210> 100
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 100
gtcagattgg cttgctcgga attg 24

<210> 101
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 101
ccagccaa 8

<210> 102
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 102
ggcataagga aatcgaaggt c 21

<210> 103
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 103
catgtcctca agtcaagaac aag 23

<210> 104
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 104
gctgggtgaa tggagcgagc agcg 24

<210> 105
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 105
tcttcgag 8

<210> 106
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 106
tacatgaagc aactccagtc cc 22

<210> 107
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 107
atcaaattcc agcaccgagc gc 22

<210> 108
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 108
gtcctggagt gacccctggc cttc 24

<210> 109
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 109
tccccgct 8

<210> 110
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 110
tacatgaagc aactccagtc cc 22

<210> 111
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 111
atcaaattcc agcaccgagc gc 22

<210> 112
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 112
gatcctggag tgacccctgg cctt 24

<210> 113
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 113
ctccccgc 8

<210> 114
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 114
tacatgaagc aactccagtc cc 22

<210> 115
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 115
atcaaattcc agcaccgagc gc 22

<210> 116
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 116
gtgtgtccct ctccccaccc gtcc 24

<210> 117

<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 117
ctgtccgg 8

<210> 118
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 118
tacatgaagc aactccagtc cc 22

<210> 119
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 119
atcaaattcc agcaccgagc gc 22

<210> 120
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 120
gttggagcgg ggagaaggcc aggg 24

<210> 121
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 121
gtcactcc 8

<210> 122
<211> 22

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 122
tacatgaagc aactccagtc cc 22

<210> 123
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 123
atcaaattcc agcaccgagc gc 22

<210> 124
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 124
gcgttggagc ggggagaagg ccag 24

<210> 125
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 125
gggtcact 8

<210> 126
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 126
tacatgaagc aactccagtc cc 22

<210> 127
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 127
atcaaattcc agcaccgagc gc 22

<210> 128
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 128
gtaccctcca ataatttggc tggc 24

<210> 129
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 129
aattccga 8

<210> 130
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 130
ggcataagga aatcgaaggt c 21

<210> 131
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 131
catgtcctca agtcaagaac aag 23

<210> 132
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 132
gataatttgg ctggcaattc cgag 24

<210> 133
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 133
caagccaa 8

<210> 134
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 134
ggcataagga aatcgaaggt c 21

<210> 135
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 135
catgtcctca agtcaagaac aag 23

<210> 136
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 136
gcaggggcca ggtgtccttc tctg 24

<210> 137
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 137
ggggcctc 8

<210> 138
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 138
cacgggcagc atgggaatag t 21

<210> 139
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 139
gctaggggag agtcccactg tcca 24

<210> 140
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 140
gaatggcagg cggaggttgt actg 24

<210> 141
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 141
ggggccag 8

<210> 142
<211> 21
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 142
ctgtgtggct ttgctttggt c 21

<210> 143
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 143
gtagggtgtg atgggaggct aagc 24

<210> 144
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 144
gagtgagaga gtgagagaga gaca 24

<210> 145
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 145
cgggccag 8

<210> 146
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 146
ctgtgtggct ttgctttggt c 21

<210> 147
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 147
gtagggtgtg atgggaggct aagc 24

<210> 148
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 148
gtgagcaggc acctgtgcca acat 24

<210> 149
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 149
gggcccgc 8

<210> 150
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 150
ctgtgtggct ttgctttggt c 21

<210> 151
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 151
gtagggtgtg atgggaggct aagc 24

<210> 152
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 152
gcgtgggggc tccgtgcccc acgc 24

<210> 153
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 153
gggtccat 8

<210> 154
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 154
gaggaagagt agctcgccga g 21

<210> 155
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 155
agaccgagtg gcagtgacag caag 24

<210> 156
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 156
gcatgggcag gggctggggt gcac 24

<210> 157
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 157

aggcccag 8

<210> 158
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 158
gggagaggga agtgtgggga ag 22

<210> 159
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 159
gtcttcctgc tctgtgcgca cgac 24

<210> 160
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 160
gaaaattgtg atttccagat ccac 24

<210> 161
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 161
aagcccaa 8

<210> 162
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 162
ttgggggctc taagttatgt a 21

<210> 163
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 163
cttcatctgt atcttcagga tca 23

<210> 164
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 164
gagcagaaaa aattgtgatt tcc 23

<210> 165
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 165
agatccac 8

<210> 166
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 166
ttgggggctc taagttatgt a 21

<210> 167
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 167
cttcatctgt atcttcagga tca 23

<210> 168
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (21)..(21)
<223> n is a, c, g, t or u

<400> 168
ggcaggcgga ggttgtactg ngg 23

<210> 169
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 169
ggcaggcgga ggttgtactg ggg 23

<210> 170
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 170
ggaaggcgga agttgtactg agg 23

<210> 171
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 171
ggcaggcgga ggttgtagtg ggg 23

<210> 172
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 172
aggaggcgga ggttgcactg agg 23

<210> 173
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 173
gggaggtgga ggttgtactg agg 23

<210> 174
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 174
ggcaggggga agctgtactg tgg 23

<210> 175
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 175
aggaggcgga ggttgcactg agg 23

<210> 176
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 176
aggaggcgga ggttgtactg agc 23

<210> 177
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 177
gggaggcgga ggttgtaatg agg 23

<210> 178
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 178
ggcaagagga ggttggactg ggg 23

<210> 179
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 179
aggaggcgga ggttgcactg agg 23

<210> 180
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 180
gggaggcgga ggttgtagtg agg 23

<210> 181
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 181
aggaagcgga ggttgtaatg agg 23

<210> 182
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 182
aggaggcgga ggttgtaatg agg 23

<210> 183
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 183
gggaggcgga ggttgtagtg agg 23

<210> 184
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 184
tccaggtgga ggctgtactg agg 23

<210> 185
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 185
aggaggcaga ggttgcactg ggg 23

<210> 186
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 186
gggaggcgga ggatgtaatg agg 23

<210> 187
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 187

cacaggcaga ggttgtactg agc 23

<210> 188
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 188
aggaggcgga ggttgtagtg agg 23

<210> 189
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 189
gggaggcgga ggttgtagtg agg 23

<210> 190
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 190
aggaggcaga ggttgaactg agg 23

<210> 191
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 191
ggcaagggga agttgtactg tgg 23

<210> 192
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 192
gggaggcaga ggttgcactg agg 23

<210> 193
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 193
gagaggcgga ggttgcactg agg 23

<210> 194
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 194
aggaggcgga ggttgcactg agg 23

<210> 195
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 195
aggaggcaga ggttgcactg agg 23

<210> 196
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 196
gggaggcaga ggttgcactg agg 23

<210> 197
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 197
cagaggcgga ggttgtagtg agg 23

<210> 198
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 198
agcaggtaga ggttggactg agg 23

<210> 199
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 199
aggaggcgga ggttgtagtg agg 23

<210> 200
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 200
gggaggcaga ggttgcactg agg 23

<210> 201
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 201
aggaggcgga ggttgcactg agg 23

<210> 202
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 202
tggaggcgga ggttgtactg agc 23

<210> 203
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 203
aggaggcaga ggttgcactg agg 23

<210> 204
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 204
gggaggcgga ggttgtagtg agg 23

<210> 205
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 205
aggaggcaga ggttgtactg agc 23

<210> 206
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 206
aggaggcgga ggttgtagtg agg 23

<210> 207
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 207
gggaggcgga ggttgtactg agc 23

<210> 208

<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 208
gggaggtgga ggttgcactg agg 23

<210> 209
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 209
gggaggcgga ggttgtagtg agg 23

<210> 210
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 210
aggaggcgga ggttgtagtg agg 23

<210> 211
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 211
gggaggcgga ggttgtagtg agg 23

<210> 212
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 212
aggaggcgga ggttgtactg agc 23

<210> 213
<211> 23

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 213
agaaggcgga ggttgtagtg agg 23

<210> 214
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 214
aggaggcgga ggttgtagtg agg 23

<210> 215
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 215
aggaggcgga ggctgcactg agg 23

<210> 216
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 216
gggaggcgga ggttgtagtg agg 23

<210> 217
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 217
gggaggcgga ggttgcactg agg 23

<210> 218
<211> 23
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 218
gggaggcaga ggttgcactg agg 23

<210> 219
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 219
aggaggcaga ggttgtaatg agg 23

<210> 220
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 220
gggaggcgga ggttgtagtg agg 23

<210> 221
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 221
ggaaggcgga ggttgtagtg agg 23

<210> 222
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 222
aggaggcaga ggttgtactg agc 23

<210> 223
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 223
gggaggcaga ggttgtactg agc 23

<210> 224
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 224
ggaaggtgaa ggctgtactg cgg 23

<210> 225
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 225
agaaggcaga ggttgcactg agg 23

<210> 226
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 226
agtaggcaga ggttgcactg agg 23

<210> 227
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 227
gggaggcaga ggttgtactg agc 23

<210> 228
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 228
gggaggcgga ggttgtactg agc 23

<210> 229
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 229
aggaggcaga ggttgtaatg agg 23

<210> 230
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 230
aggaggcgga ggttgtagtg agg 23

<210> 231
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 231
gggaggcaga ggttgtactg agc 23

<210> 232
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 232
gccaggcggg tgctgtactg ggg 23

<210> 233
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 233
aggaggcgga ggttgtactg ggc 23

<210> 234
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 234
tggaggcgga ggttgtactg agc 23

<210> 235
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 235
gggaggtgga ggttgtactg agc 23

<210> 236
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 236
aggaggtgga ggttgtaatg agg 23

<210> 237
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 237
aggaggcgga ggttgtagtg agg 23

<210> 238
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 238
gggaggcaga ggttgtactg agc 23

<210> 239
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 239
aggaggcaga ggttgcactg agg 23

<210> 240
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 240
gggaggcaga ggttgtagtg agg 23

<210> 241
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 241
gggaggcgga ggttgtactg agc 23

<210> 242
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 242
ggtaggcaaa ggttgtacca ggg 23

<210> 243
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 243
gggaggcgga ggttgtagtg agg 23

<210> 244
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 244
aggaggcaga ggttgtactg agc 23

<210> 245
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 245
gggaggcaga ggttgtagtg agg 23

<210> 246
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 246
gggaggcgga ggttgtagtg agg 23

<210> 247
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 247
gggaggcaga ggttgtagtg agg 23

<210> 248
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 248

gggaggcgga ggttgtactg agc 23

<210> 249
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 249
gggaggtgga ggttgcactg agg 23

<210> 250
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 250
cagaggcaga ggttgtactg agc 23

<210> 251
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 251
gggaggcaga ggttgtactg agt 23

<210> 252
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 252
gggaggcaga ggttgtactg agc 23

<210> 253
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 253
agaaggcgga ggttgtagtg agg 23

<210> 254
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 254
cgtctgcgag ggtactagtg aga 23

<210> 255
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 255
gggaggcaga ggttgtactg agc 23

<210> 256
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 256
gggagacgga ggttgtagtg agg 23

<210> 257
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 257
aggaggcaga ggttgtactg agc 23

<210> 258
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 258
aggaggcgga ggttgtactg agc 23

<210> 259
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 259
agaaggcaga ggttgtactg agc 23

<210> 260
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 260
gccaggctga ggatgtactg tgg 23

<210> 261
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 261
aggaggcgga ggttgtactg agc 23

<210> 262
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 262
gggaggcaga ggttgtagtg agg 23

<210> 263
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 263
agaa 4

<210> 264
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 264
tctc 4

<210> 265
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 265
agac 4

<210> 266
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 266
agaa 4

<210> 267
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 267
agac 4

<210> 268
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 268
cagg 4

<210> 269

<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 269
atag 4

<210> 270
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 270
agac 4

<210> 271
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 271
agcc 4

<210> 272
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 272
agaa 4

<210> 273
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 273
atac 4

<210> 274
<211> 4

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 274
aaac 4

<210> 275
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 275
atac 4

<210> 276
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 276
agaa 4

<210> 277
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 277
cgtc 4

<210> 278
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 278
agta 4

<210> 279
<211> 4
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 279
agac 4

<210> 280
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 280
taaca 5

<210> 281
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 281
taaca 5

<210> 282
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 282
ctctagacg 9

<210> 283
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (21)..(21)
<223> n is a, c, g, or t

<400> 283
gcaggcacct gtgccaacat ngg 23

<210> 284
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 284
gcaggcacct gtgccaacat ggg 23

<210> 285
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 285
acaggcactg atgccaactt tgg 23

<210> 286
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 286
taatgccctg gagcctccct ggc 23

<210> 287
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 287
gcagggcgcg ccgagagcag cgg 23

<210> 288
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 288
ccagccaccc agcccctcct ccc 23

<210> 289
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 289
gtaagcatat gatagtccat ttt 23

<210> 290
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 290
ccgcgtccct gcgcaaaccc agg 23

<210> 291
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 291
gtgcacccct gctcctaccc ccc 23

<210> 292
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 292
ccagggagca atggcagcgc gcc 23

<210> 293
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 293
ggcggaagtt gtactgaggt gag 23

<210> 294
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 294
gcaggaactg gagtgcacag gtg 23

<210> 295
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 295
atgat 5

<210> 296
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 296
tatctgatcc 10

<210> 297
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 297
gcggccaggg 10

<210> 298
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 298
cccagcctc 9

<210> 299

<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 299
tataatagtc 10

<210> 300
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 300
cgctccacc 9

<210> 301
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 301
tgcaccttcc 10

<210> 302
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 302
cggaagggc 9

<210> 303
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 303
gcagtatggg 10

<210> 304
<211> 8

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 304
atgatgcg 8

<210> 305
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (21)..(21)
<223> n is a, c, g, or t

<400> 305
gggcaggggc tggggtgcac ngg 23

<210> 306
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 306
cggcaggggc tgaggggcac tgg 23

<210> 307
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 307
gggtaggagc aggggtgcac tgg 23

<210> 308
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 308

gggcaggggc tggggtgcac agg 23

<210> 309
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 309
gggcaggaac tggagtgcac agg 23

<210> 310
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 310
gggcaggaac tggagtgcac agg 23

<210> 311
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 311
cagcaggggc tggggtgcac agg 23

<210> 312
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 312
gggaagggcc tggggtacac ggg 23

<210> 313
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 313
gggccggggc aggggtgcac agg 23

<210> 314
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 314
agacaggggc cggggtgcac agg 23

<210> 315
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 315
gggcaggaac tggagtgcac cgg 23

<210> 316
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 316
aggcaggaac tggagtgcac ggg 23

<210> 317
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 317
aggcaggaac tggagtgcac agg 23

<210> 318
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 318
aggaagggac tggggtgcac tgg 23

<210> 319
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 319
aggcaggaac tggagtgcac agg 23

<210> 320
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 320
aggtgggggc tggggtgcac agg 23

<210> 321
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 321
aggcaggaac tggggtgcac ggg 23

<210> 322
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 322
tggcaggggc aggggtgaac tgg 23

<210> 323
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 323
gggcaggaac tggagtgcac ggg 23

<210> 324
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 324
ggccaggggc tggggagcac agg 23

<210> 325
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 325
gggcagggct gggggtgcac agg 23

<210> 326
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 326
tgggtggggc tggggtgcac tgg 23

<210> 327
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 327
gggagggggc tggggagcac agg 23

<210> 328
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 328
gggcaggaac tggagtacac ggg 23

<210> 329

<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 329
gagaaggagc tggggagcac tgg 23

<210> 330
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 330
gggcaggaac tggagtgcac cag 23

<210> 331
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 331
gggcaagggc aggggtgcac cag 23

<210> 332
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 332
aagaaggggc aagggtgcac agg 23

<210> 333
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 333
ggccaggagc aggggtgcac ggg 23

<210> 334
<211> 23

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 334
tggcagcggc tggggagcac tgg 23

<210> 335
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 335
gggcgtgggc aggggtgcac tgg 23

<210> 336
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 336
gggcagtggc tggggtgcat tgg 23

<210> 337
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 337
ggccaggagc tggggtgctc agg 23

<210> 338
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 338
cctcaggggc tggggtgaac agg 23

<210> 339
<211> 23
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 339
tggcagggtc tggggtgcac aga 23

<210> 340
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 340
gagcagggtc tggggtgcat ggg 23

<210> 341
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 341
gagcagggac tgaggggcac agg 23

<210> 342
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 342
gagcaggggc tggggggcac tgg 23

<210> 343
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 343
tggcaggggt aagggtgcac tgg 23

<210> 344

<400> 344
000

<210> 345
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 345
agacagaggc tggagtgcac tgg 23

<210> 346
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 346
aggcaggggc tggagttcac agg 23

<210> 347
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 347
aggaagggac cagggtgcac cag 23

<210> 348
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 348
ggccaggagc aggggtgcac agg 23

<210> 349
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 349
gggccggggc tggggtgcca ggg 23

<210> 350
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 350
gggcgggggc tggggagcac agg 23

<210> 351
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 351
aggcaggagc cagggtgcag agg 23

<210> 352
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 352
gggcagaggc tggagtgccc agg 23

<210> 353
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 353
ggacaggggc aggggtgccc ggg 23

<210> 354
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 354
agggaggggc tggggtgcac gga 23

<210> 355

<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 355
gggcaggaac tggagtgcat agg 23

<210> 356
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 356
aggcaggaac tggagtgcac aag 23

<210> 357
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 357
gggcagaggc tagggtgcag tgg 23

<210> 358
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 358
aggtaggggt tgggggggcac agg 23

<210> 359
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 359
gggcagaagc aggggtgctc agg 23

<210> 360
<211> 23

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 360
ggggaggggt gggggtgcac cgg 23

<210> 361
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 361
gagcaggggc tgggggggcac tgg 23

<210> 362
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 362
gggcagaggc tggagtgccc agg 23

<210> 363
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 363
gggtgggggc tggggtgccc agg 23

<210> 364
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 364
gggcaagggc aggggtgccc tgg 23

<210> 365
<211> 23
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 365
gagagggagc tggggtgcac ggg 23

<210> 366
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 366
aggcagggac tgaggtgcat agg 23

<210> 367
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 367
gggccagggc tgaggtgcac agg 23

<210> 368
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 368
tgggaggggc tagagtgcac agg 23

<210> 369
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 369
ccgcaggggc tgggatgctg ggg 23

<210> 370

<400> 370
000

<210> 371
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 371
gaggaggggc tggggtgccc tgg 23

<210> 372
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 372
gggcaaaggc cggggtgccc agg 23

<210> 373
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 373
aggcgggggc tgggggggctc ggg 23

<210> 374
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 374
aggcaggggc cagggtccac agg 23

<210> 375

<400> 375
000

<210> 376
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 376
gggttggggt tggggtgcac agg 23

<210> 377
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 377
aggcaggggc cggggtgcgc agg 23

<210> 378
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 378
gggcacagac tggggtgcat tgg 23

<210> 379
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 379
gggctggggc tgaggtgcgc cgg 23

<210> 380
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 380
aggcaggggc tgggggggcaa ggg 23

<210> 381
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 381
gagcgggagc tgggggggcac agg 23

<210> 382
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 382
gggcagggac tggggtgctt agg 23

<210> 383
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 383
gggaaggggc tggagggcac agg 23

<210> 384
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 384
gggcagggga aggggtggac tgg 23

<210> 385
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 385
agacaggggc tggagtgcag tgg 23

<210> 386
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 386

gggcagaggc tggagtgcaa tgg 23

<210> 387
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 387
gggctggggc tggggagcag ggg 23

<210> 388
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 388
aggaaagggc tggagtgcag ggg 23

<210> 389
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 389
gggcaggaac tggagtgcac cag 23

<210> 390
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 390
aggcaggaac tggagtgcac aag 23

<210> 391
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 391
aggcagagcc tggggtgcag ggg 23

<210> 392
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 392
gggcagggcc aggggagcac agg 23

<210> 393
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 393
agccaggggc tgggggggaac agg 23

<210> 394
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 394
gggcagggga tggggtgcag tgg 23

<210> 395
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 395
aggcaaggcc tggggtgccc agg 23

<210> 396
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 396
gggctggggc tggggagcac ggg 23

<210> 397
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 397
gagaaggggc tgggaagcac agg 23

<210> 398
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 398
aggcaggaac tggagtgcac aag 23

<210> 399
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 399
ggggaggggc tggggtgcca ggg 23

<210> 400
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 400
aggaaggggc tggggaaaac agg 23

<210> 401
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 401
ccccaggggc tggggtgcct ggg 23

<210> 402
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 402
aagcagaggc tgaagtgcac agg 23

<210> 403
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 403
aggcaggaac tagagtgcac agg 23

<210> 404
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 404
gggcaggggg tggggtccac agg 23

<210> 405

<400> 405
000

<210> 406
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 406
ggggaggggc tggggagcac gga 23

<210> 407
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 407
aggcagaggc tggagtggac cgg 23

<210> 408
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 408
gggtaggggc tgggggatac cgg 23

<210> 409
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 409
gggaagggtc tggagtccac tgg 23

<210> 410
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 410
gggcaggaac tagagtgcac ggg 23

<210> 411
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 411
gggcagggac tggggtgctc tgg 23

<210> 412
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 412

gagtaggggc aggggtgctc tgg 23

<210> 413
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 413
aggaagggcc tggggtgcac aga 23

<210> 414
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 414
ggccaggggc tggggtgcac ggt 23

<210> 415
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 415
aggcaggggc cagggtgcat ggg 23

<210> 416
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 416
gggcagagga tggggtgcag ggg 23

<210> 417
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 417
cggcaggggc tggagtgcag tgg 23

<210> 418
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 418
aggcaggatc tggagtgcac agg 23

<210> 419
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 419
agacaggagc tggagtgcac aag 23

<210> 420
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 420
tggcaggggc agggatgctc tgg 23

<210> 421
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 421
cctcaggggt tgggatgcac tgg 23

<210> 422
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 422
gagcagggtc aggggtgcag agg 23

<210> 423
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 423
caggagtggc tggggtgcac agg 23

<210> 424
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 424
gggcctgggc tgagatgcac ggg 23

<210> 425
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 425
acgcaggggc tagggagcac aag 23

<210> 426
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 426
gggctggggc tggggaggac ggg 23

<210> 427
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 427
gggcagggat tggggggcac agg 23

<210> 428
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 428
agggaggggc cgggctgcac tgg 23

<210> 429
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 429
aaaa 4

<210> 430
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 430
aaaa 4

<210> 431
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 431
aaaa 4

<210> 432
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 432
aaaa 4

<210> 433

<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 433
aaaa 4

<210> 434
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 434
aaaaa 5

<210> 435
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 435
ccta 4

<210> 436
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 436
ccta 4

<210> 437
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 437
ttaa 4

<210> 438
<211> 4

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 438
aaaa 4

<210> 439
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 439
aaaca 5

<210> 440
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 440
aacag 5

<210> 441
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 441
aaat 4

<210> 442
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 442
aaaa 4

<210> 443
<211> 4
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 443
attg 4

<210> 444
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 444
aaat 4

<210> 445
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 445
aaga 4

<210> 446
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 446
aaat 4

<210> 447
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 447
tgaa 4

<210> 448
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 448
aacc 4

<210> 449
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 449
aggt 4

<210> 450
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 450
acac 4

<210> 451
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 451
caat 4

<210> 452
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 452
agag 4

<210> 453
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 453
aaag 4

<210> 454
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 454
aaaag 5

<210> 455
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 455
aaaa 4

<210> 456
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 456
aacg 4

<210> 457
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 457
acga 4

<210> 458
<211> 4
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 458
aacc 4

<210> 459
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 459
aaaa 4

<210> 460
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 460
aaaa 4

<210> 461
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 461
aaaaa 5

<210> 462
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 462
cccct 5

<210> 463
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 463
aaaaa 5

<210> 464
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 464
aaaaa 5

<210> 465
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 465
aaag 4

<210> 466
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 466
tgat 4

<210> 467
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 467
atac 4

<210> 468
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 468
aaaa 4

<210> 469
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 469
atat 4

<210> 470
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 470
acat 4

<210> 471
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 471
aata 4

<210> 472
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 472
aaaa 4

<210> 473
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 473

taat 4

<210> 474
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 474
cctta 5

<210> 475
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 475
atag 4

<210> 476
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 476
cagt 4

<210> 477
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 477
ctaa 4

<210> 478
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 478
acaa 4

<210> 479
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 479
agcc 4

<210> 480
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (25)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (31)..(32)
<223> n is a, c, g, or t

<400> 480
gaatggcagg cggaggttgt actgnnnncc nn 32

<210> 481
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 481
gaatggcagg cggaggttgt actgggggcc ag 32

<210> 482
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 482
aaacggaagc cgcacgtctc actagtaccc tc 32

<210> 483
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 483
acacccctca 10

<210> 484
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 484
ggggccag 8

<210> 485
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 485
gtaccctc 8

<210> 486
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (25)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (31)..(32)
<223> n is a, c, g, or t

<400> 486
gtgagcaggc acctgtgcca acatnnnncc nn 32

<210> 487
<211> 32

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 487
gtgagcaggc acctgtgcca acatgggccc gc 32

<210> 488
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 488
gggcccgc 8

<210> 489
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (25)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (31)..(32)
<223> n is a, c, g, or t

<400> 489
gcatgggcag gggctggggt gcacnnnncc nn 32

<210> 490
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 490
gcatgggcag gggctggggt gcacaggccc ag 32

<210> 491
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 491
gcaggaagcg tcgccggggg gcccacaagg gt 32

<210> 492
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 492
gaagctccgc 10

<210> 493
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 493
aggcccag 8

<210> 494
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 494
acaagggt 8

<210> 495
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (25)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (31)..(32)

<223> n is a, c, g, or t

<400> 495
tgaggaccgc cctgggcctg ggagnnnncc nn 32

<210> 496
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 496
tgaggaccgc cctgggcctg ggagaatccc tt 32

<210> 497
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 497
gaaggaccac cctaggcctg ggagactccc tc 32

<210> 498
<211> 4
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 498
gaaa 4

<210> 499
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 499
aatccctt 8

<210> 500
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 500
actccctc 8

<210> 501
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (25)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (31)..(32)
<223> n is a, c, g, or t

<400> 501
ggcctggctg atgaggccgc acatnnnncc nn 32

<210> 502
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 502
ggcctggctg atgaggccgc acatgtggcc ac 32

<210> 503
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 503
gtggccac 8

<210> 504
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 504

gtcacctgcc tcgtggaata cgg 23

<210> 505
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 505
gcacctgcct cgtggaatac ggt 23

<210> 506
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 506
gttcagcgtg tccggctttg gc 22

<210> 507
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 507
gtggtgagca agggcgagga gctg 24

<210> 508
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 508
gggcgaggag ctgttcaccg gggt 24

<210> 509
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 509
gtgaacttgt ggccgtttac gtcg 24

<210> 510
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 510
gcgtccagct cgaccaggat gggc 24

<210> 511
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 511
gcggtgaaca gctcctcgcc cttg 24

<210> 512
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 512
gggcaccacc ccggtgaaca gctc 24

<210> 513
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 513
ggcaccaccc cggtgaacag ctcc 24

<210> 514
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 514
gggatgggca ccaccccggt gaac 24

<210> 515
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 515
gcgtgtccgg ctttggcgag acaa 24

<210> 516
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 516
gtccggcttt ggcgagacaa atca 24

<210> 517
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 517
gatcacctgc ctcgtggaat acgg 24

<210> 518
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 518
gacgctgaac ttgtggccgt ttac 24

<210> 519
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 519
gccaaagccg gacacgctga actt 24

<210> 520
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 520
ggaacttgtg gccgtttacg tcg 23

<210> 521
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 521
gaacttgtgg ccgtttacgt cg 22

<210> 522
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 522
gacttgtggc cgtttacgtc g 21

<210> 523
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 523
gcttgtggcc gtttacgtcg 20

<210> 524
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 524
gttgtggccg tttacgtcg 19

<210> 525

<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 525
gtcacctgcc tcgtggaata cgg 23

<210> 526
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 526
gcacctgcct cgtggaatac gg 22

<210> 527
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 527
gacctgcctc gtggaatacg g 21

<210> 528
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 528
gcctgcctcg tggaatacgg 20

<210> 529
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 529
ggttctgggt acttttatct gtcc 24

<210> 530
<211> 24

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 530
gtctgcctaa caggaggtgg gggt 24

<210> 531
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 531
gaatatcagg agactaggaa ggag 24

<210> 532
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 532
gcctccctgc agggctgctc cc 22

<210> 533
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 533
gagctagtct tcttcctcca accc 24

<210> 534
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 534
gatctgtccc ctccacccca cagt 24

<210> 535
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 535
ggcccaaatg aaaggagtga gagg 24

<210> 536
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 536
gcatcctctt gctttctttg cctg 24

<210> 537
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 537
ggagtcgcca gaggccggtg gtgg 24

<210> 538
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 538
gcccagcggc cggatatcag ctgc 24

<210> 539
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 539
ggaagggaac atattactat tgc 23

<210> 540
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 540
gtggagtggc ctgctatcag ctac 24

<210> 541
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 541
gaggaaggga acatattact attg 24

<210> 542
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 542
gtgaattctc atcagctaaa atgc 24

<210> 543
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 543
gctcactcac ccacacagac acac 24

<210> 544
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 544
ggaagaattt cattctgttc tcag 24

<210> 545
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 545
gctcagtttt cctggattat gcct 24

<210> 546
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 546
gcgttggagc ggggagaagg ccag 24

<210> 547
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 547
gggccgcgga gatagctgca gggc 24

<210> 548
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 548
gcccacccgg cggcgcctcc ctgc 24

<210> 549
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 549
gcgtggcagc tgatatccgg ccgc 24

<210> 550
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 550
gccgcggcgc gacgtggagc cagc 24

<210> 551
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 551
gtgctcccca gcccaaaccg ccgc 24

<210> 552
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 552
gtcagattgg cttgctcgga attg 24

<210> 553
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 553
gctgggtgaa tggagcgagc agcg 24

<210> 554
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 554
gtcctggagt gacccctggc cttc 24

<210> 555
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 555
gatcctggag tgacccctgg cctt 24

<210> 556
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 556
gtgtgtccct ctccccaccc gtcc 24

<210> 557
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 557
gttggagcgg ggagaaggcc aggg 24

<210> 558
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 558
gcgttggagc ggggagaagg ccag 24

<210> 559
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 559
gtaccctcca ataatttggc tggc 24

<210> 560
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 560
gataatttgg ctggcaattc cgag 24

<210> 561
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 561
gaaaattgtg atttccagat ccac 24

<210> 562
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 562
gagcagaaaa aattgtgatt tcc 23

<210> 563
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 563
gcaggggcca ggtgtccttc tctg 24

<210> 564
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 564
gaatggcagg cggaggttgt actg 24

<210> 565
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 565

gagtgagaga gtgagagaga gaca 24

<210> 566
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 566
gtgagcaggc acctgtgcca acat 24

<210> 567
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 567
gcgtgggggc tccgtgcccc acgc 24

<210> 568
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 568
gcatgggcag gggctggggt gcac 24

<210> 569
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 569
gggccaggtg tccttctctg 20

<210> 570
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 570
ggcaggcgga ggttgtactg 20

<210> 571
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 571
gagagagtga gagagagaca 20

<210> 572
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 572
gcaggcacct gtgccaacat 20

<210> 573
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 573
gggggctccg tgccccacgc 20

<210> 574
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 574
gggcaggggc tggggtgcac 20

<210> 575
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 575
tggcttagca cctctccat 19

<210> 576
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 576
agaggagcct tctgactgct gcaga 25

<210> 577
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 577
tccgtcttcc tccactcc 18

<210> 578
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 578
tagagaactg ggtagtgtg 19

<210> 579
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 579
gtacatgaag caactccagt ccca 24

<210> 580
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 580
tggtgattat gggagaactg gagc 24

<210> 581
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 581
ggcataagga aatcgaaggt c 21

<210> 582
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 582
acacgggcag catgggaata gtc 23

<210> 583
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 583
cctgtgtggc tttgctttgg tcg 23

<210> 584
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 584
ggaggaagag tagctcgccg agg 23

<210> 585
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 585
agggagaggg aagtgtgggg aagg 24

<210> 586

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 586
agaactcagg accaacttat tctg 24

<210> 587
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 587
atgacagaca caaccagagg gca 23

<210> 588
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 588
taggaaggag gaggcctaag 20

<210> 589
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 589
ccaatattgc atgggatgg 19

<210> 590
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 590
atcaaattcc agcaccgagc gc 22

<210> 591
<211> 23

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 591
accattgagg acgtttgtct cac 23

<210> 592
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 592
catgtcctca agtcaagaac aag 23

<210> 593
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 593
gctaggggag agtcccactg tcca 24

<210> 594
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 594
gtagggtgtg atgggaggct aagc 24

<210> 595
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 595
agaccgagtg gcagtgacag caag 24

<210> 596
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 596
gtcttcctgc tctgtgcgca cgac 24

<210> 597
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (35)..(44)
<223> n is a, c, g, or t

<400> 597
tagcggccgc tcatgcgcgg cgcattacct ttacnnnnnn nnnnggatcc tctagagtcg 60

<210> 598
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 598
acaggaaaca gctatgacca tgaaagcttg catgcctgca ggtcgactct agaggatc 58

<210> 599
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 599
ctacacgacg ctcttccgat ctcctggagc gtgtacgttg g 41

<210> 600
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 600
ctacacgacg ctcttccgat ctcctgtggt cccagctact tg 42

<210> 601
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 601
ctacacgacg ctcttccgat ctatctgcga tgtcctcgag g 41

<210> 602
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 602
ctacacgacg ctcttccgat cttggtgtgc gcctctaacg 40

<210> 603
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 603
ctacacgacg ctcttccgat ctggagtctt gctttgtcac tcaga 45

<210> 604
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 604
ctacacgacg ctcttccgat ctagcctaga cccagtccca t 41

<210> 605
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 605
ctacacgacg ctcttccgat ctgctgggca tagtagtgga ct 42

<210> 606
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 606
ctacacgacg ctcttccgat cttggggagg ctgagacacg a 41

<210> 607

<400> 607
000

<210> 608
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 608
ctacacgacg ctcttccgat ctcttgggag gctgaggcaa g 41

<210> 609
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 609
agacgtgtgc tcttccgatc tcaggaggat gagagccagg 40

<210> 610
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 610
agacgtgtgc tcttccgatc tcagggtctc actctatcac cca 43

<210> 611
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 611
agacgtgtgc tcttccgatc tactgaatgg gttgaacttg gc 42

<210> 612
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 612
agacgtgtgc tcttccgatc tgagacagaa tcttgctctg tctcc 45

<210> 613
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 613
agacgtgtgc tcttccgatc ttcccagcta cttgggaggc 40

<210> 614
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 614
agacgtgtgc tcttccgatc tcctgcccaa atagggaagc ag 42

<210> 615
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 615
agacgtgtgc tcttccgatc ttggcgcctt agtctctgct ac 42

<210> 616
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 616
agacgtgtgc tcttccgatc tgcatgagac acagtttcac tctg 44

<210> 617
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 617
agacgtgtgc tcttccgatc tgagagagtc tcactgcgtt gc 42

<210> 618
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 618
ctacacgacg ctcttccgat cttctctcac ccactgggca c 41

<210> 619
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 619
agacgtgtgc tcttccgatc tgcttccaga cgagtgcaga 40

<210> 620
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 620
ctacacgacg ctcttccgat ctaagttttc aaaccagaag aactacgac 49

<210> 621
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 621

ctacacgacg ctcttccgat ctccggtata agtcctggag cg 42

<210> 622
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 622
ctacacgacg ctcttccgat ctgccaggga gcaatggcag 40

<210> 623
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 623
ctacacgacg ctcttccgat ctcctcgaat tccacggggt t 41

<210> 624
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 624
ctacacgacg ctcttccgat ctgttggtgg gagggaagtg ag 42

<210> 625
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 625
ctacacgacg ctcttccgat ctgatggcgg ttgtagcggc 40

<210> 626
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 626
ctacacgacg ctcttccgat ctcacataaa cctatgtttc agcaga 46

<210> 627
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 627
ctacacgacg ctcttccgat ctgctagttg gattgaagca gggt 44

<210> 628
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 628
ctacacgacg ctcttccgat ctttgagtgc ggcagcttcc 40

<210> 629
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 629
ctacacgacg ctcttccgat ctataaccct cccaggcaaa gtc 43

<210> 630
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 630
ctacacgacg ctcttccgat ctagcctgca catctgagct c 41

<210> 631
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 631
ctacacgacg ctcttccgat ctggagcatt gaagtgcctg g 41

<210> 632
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 632
agacgtgtgc tcttccgatc tcagcctggg accactga 38

<210> 633
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 633
agacgtgtgc tcttccgatc tcatcctcga cagtcgcgg 39

<210> 634
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 634
agacgtgtgc tcttccgatc tgactgatca agtagaatac tcatggg 47

<210> 635
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 635
agacgtgtgc tcttccgatc tccctgccag cactgaagc 39

<210> 636
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 636
agacgtgtgc tcttccgatc tggttcctat ctttctagac caggagt 47

<210> 637
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 637
agacgtgtgc tcttccgatc tagtgtggag ggctcaggg 39

<210> 638
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 638
agacgtgtgc tcttccgatc tgatgggcag aggaaggcaa 40

<210> 639
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 639
agacgtgtgc tcttccgatc ttcactctca tgagcgtccc a 41

<210> 640
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 640
ctacacgacg ctcttccgat ctaaggttcc ttgcggttcg c 41

<210> 641
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 641
agacgtgtgc tcttccgatc tcgctgccat tgctccct 38

<210> 642

<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 642
ctacacgacg ctcttccgat cttctcgcac attcttcacg tcc 43

<210> 643
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 643
agacgtgtgc tcttccgatc taggaacctt cccgacttag gg 42

<210> 644
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 644
ctacacgacg ctcttccgat ctcccgccca tcttctagaa agac 44

<210> 645
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 645
ctacacgacg ctcttccgat cttgccaggt gagggactgg 40

<210> 646
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 646
agacgtgtgc tcttccgatc ttctgggagt tctctgctgc c 41

<210> 647
<211> 41

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 647
agacgtgtgc tcttccgatc ttgcccaacc ttagcaagga g 41

<210> 648
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 648
ctacacgacg ctcttccgat cttaccttgg agcaacggcg 40

<210> 649
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 649
agacgtgtgc tcttccgatc tcccaggacg aggatggag 39

<210> 650
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 650
gatagtcact ccaggggttg 20

<210> 651
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 651
gtggtgaacc aatcagtcct 20

<210> 652
<211> 60
<212> PRT

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 652

Met Ala Ala Phe Lys Pro Asn Ser Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met
50 55 60

<210> 653
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 653

Ala Arg Arg Leu Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala
1 5 10 15

His Arg Leu Leu Arg Thr Arg Arg Leu Leu Lys Arg Glu Gly Val Leu
20 25 30

Gln Ala Ala Asn Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn
35 40 45

Thr Pro Trp Gln Leu Arg Ala Ala Ala Leu Asp Arg
50 55 60

<210> 654
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 654

Ala Arg Arg Leu Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala
1 5 10 15

His Arg Leu Leu Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu
20 25 30

Gln Ala Ala Asp Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn
35 40 45

Thr Pro Trp Gln Leu Arg Ala Ala Ala Leu Asp Arg
50 55 60

<210> 655
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 655

Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu His Leu Ile Lys
1 5 10 15

His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu Gly Glu Thr Ala Asp
20 25 30

Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Gly Asn Ala His Ala
35 40 45

Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu
50 55 60

<210> 656
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 656

Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu His Leu Ile Lys
1 5 10 15

His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu Gly Glu Thr Ala Asp
20 25 30

Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Asn Asn Ala His Ala
35 40 45

Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu
50 55 60

<210> 657
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 657

Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln Arg
1 5 10 15

Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu
20 25 30

Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser
35 40 45

Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
50 55 60

<210> 658
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 658

Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln Arg
1 5 10 15

Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu
20 25 30

Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser
35 40 45

Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
50 55 60

<210> 659
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 659

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

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
20 25 30

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
35 40 45

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr
50 55 60

<210> 660
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 660

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

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
20 25 30

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
35 40 45

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr
50 55 60

<210> 661
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 661

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

Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe
20 25 30

Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met
35 40 45

Glu Met Lys Ala Tyr His Ala Ile Ser Arg Ala Leu
50 55 60

<210> 662
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 662

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

Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe
20 25 30

Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met
35 40 45

Glu Met Lys Ala Tyr His Ala Ile Ser Arg Ala Leu
50 55 60

<210> 663
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 663

Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn Leu Ser Pro
1 5 10 15

Glu Leu Gln Asp Glu Ile Gly Thr Ala Phe Ser Leu Phe Lys Thr Asp
20 25 30

Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Ile Gln Pro Glu Ile Leu
35 40 45

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

<210> 664
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 664

Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn Leu Ser Ser
1 5 10 15

Glu Leu Gln Asp Glu Ile Gly Thr Ala Phe Ser Leu Phe Lys Thr Asp
20 25 30

Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val Gln Pro Glu Ile Leu
35 40 45

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

<210> 665
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 665

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

Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp His
20 25 30

Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile Pro
35 40 45

Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
50 55 60

<210> 666
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 666

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

Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp His
20 25 30

Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile Pro
35 40 45

Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
50 55 60

<210> 667
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 667

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
1 5 10 15

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
20 25 30

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
35 40 45

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr
50 55 60

<210> 668
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 668

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
1 5 10 15

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
20 25 30

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
35 40 45

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr
50 55 60

<210> 669
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 669

Phe Pro Asn Phe Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu
1 5 10 15

Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu
20 25 30

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

Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe
50 55 60

<210> 670
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 670

Phe Pro Asn Phe Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu
1 5 10 15

Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu
20 25 30

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

Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser Phe

50 55 60

<210> 671
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 671

Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly Asn
1 5 10 15

Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu Trp
20 25 30

Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser Lys
35 40 45

Lys Gln Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp
50 55 60

<210> 672
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 672

Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly Asn
1 5 10 15

Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu Trp
20 25 30

Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser Lys
35 40 45

Lys Gln Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp
50 55 60

<210> 673
<211> 60
<212> PRT
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 673

Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe
1 5 10 15

Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr Gly Lys Gly Lys
20 25 30

Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn Leu Leu Arg Gly
35 40 45

Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
50 55 60

<210> 674
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 674

Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe
1 5 10 15

Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr Gly Lys Gly Lys
20 25 30

Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn Leu Leu Arg Gly
35 40 45

Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
50 55 60

<210> 675
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 675

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
1 5 10 15

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala

20 25 30

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
35 40 45

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala
50 55 60

<210> 676
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 676

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
1 5 10 15

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
20 25 30

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln
35 40 45

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala
50 55 60

<210> 677
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 677

Gln Glu Val Met Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu
1 5 10 15

Phe Glu Glu Ala Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu
20 25 30

Lys Leu Ser Ser Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu
35 40 45

Phe Val Ser Arg Ala Pro Asn Arg Lys Met Ser Gly
50 55 60

<210> 678
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 678

Gln Glu Val Met Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu
1 5 10 15

Phe Glu Glu Ala Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu
20 25 30

Lys Leu Ser Ser Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu
35 40 45

Phe Val Ser Arg Ala Pro Asn Arg Lys Met Ser Gly
50 55 60

<210> 679
<211> 55
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 679

Gln Gly His Met Glu Thr Val Lys Ser Ala Lys Arg Leu Asp Glu Gly
1 5 10 15

Val Ser Val Leu Arg Val Pro Leu Thr Gln Leu Lys Leu Lys Asp Leu
20 25 30

Glu Lys Met Val Asn Arg Glu Arg Glu Pro Lys Leu Tyr Glu Ala Leu
35 40 45

Lys Ala Arg Leu Glu Ala His
50 55

<210> 680
<211> 59
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 680

Ala His Lys Asp Thr Leu Arg Ser Ala Lys Arg Phe Val Lys His Asn
1 5 10 15

Glu Lys Ile Ser Val Lys Arg Val Trp Leu Thr Glu Ile Lys Leu Ala
20 25 30

Asp Leu Glu Asn Met Val Asn Tyr Lys Asn Gly Arg Glu Ile Glu Leu
35 40 45

Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala Tyr
50 55

<210> 681
<211> 55
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 681

Lys Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp
1 5 10 15

Lys Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln
20 25 30

Val Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp
35 40 45

Asn Ala Thr Met Val Arg Val
50 55

<210> 682
<211> 55
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 682

Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn Pro Phe Tyr
1 5 10 15

Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val Glu Lys Thr Gln
20 25 30

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

Asn Gly Asp Met Val Arg Val
50 55

<210> 683
<211> 55
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 683

Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr Leu Val Pro Ile Tyr Ser
1 5 10 15

Trp Gln Val Ala Lys Gly Ile Leu Pro Asp Arg Ala Val Val Gln Gly
20 25 30

Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp Asp Ser Phe Asn Phe Lys
35 40 45

Phe Ser Leu His Pro Asn Asp
50 55

<210> 684
<211> 53
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 684

Asp Val Phe Cys Lys Val Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile
1 5 10 15

Val Pro Ile Tyr Ala Trp Gln Val Ala Glu Asn Ile Leu Pro Asp Ile
20 25 30

Asp Cys Lys Gly Tyr Arg Ile Asp Asp Ser Tyr Thr Phe Cys Phe Ser
35 40 45

Leu His Lys Tyr Asp
50

<210> 685
<211> 58
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 685

Leu Val Glu Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala
1 5 10 15

Ser Cys His Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu
20 25 30

Asp His Lys Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys
35 40 45

Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile
50 55

<210> 686
<211> 56
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 686

Leu Ile Ala Phe Gln Lys Asp Glu Lys Ser Lys Val Glu Phe Ala Tyr
1 5 10 15

Tyr Ile Asn Cys Asp Ser Ser Asn Gly Arg Phe Tyr Leu Ala Trp His
20 25 30

Asp Lys Gly Ser Lys Glu Gln Gln Phe Arg Ile Ser Thr Gln Asn Leu
35 40 45

Val Leu Ile Gln Lys Tyr Gln Val
50 55

<210> 687
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 687

Asp Glu Leu Gly Lys Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro
1 5 10 15

Pro Val Arg

<210> 688
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 688

Asn Glu Leu Gly Lys Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro
1 5 10 15

Pro Val Arg

<210> 689
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 689

Met Ala Ala Phe Lys Pro Asn Ser Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu
50

<210> 690
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 690

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu
50

<210> 691
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 691

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

Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg Thr
20 25 30

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

Glu Asn
50

<210> 692
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 692

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

Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala
20 25 30

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

Glu Asn
50

<210> 693
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 693

Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala Ala
1 5 10 15

Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu
20 25 30

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

Glu Thr
50

<210> 694
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 694

Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala Ala
1 5 10 15

Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu
20 25 30

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

Glu Thr
50

<210> 695
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 695

Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Gly Asn Ala
1 5 10 15

His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala Leu
20 25 30

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

Tyr Ser
50

<210> 696
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 696

Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Asp Asn Ala
1 5 10 15

His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala Leu
20 25 30

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

Tyr Ser
50

<210> 697
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 697

His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu Ile Leu Leu Phe
1 5 10 15

Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser Gly Gly Leu Lys
20 25 30

Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg Pro Ala Leu Ser Gly
35 40 45

Asp Ala
50

<210> 698
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 698

His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu Asn Leu Leu Phe
1 5 10 15

Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser Gly Gly Leu Lys
20 25 30

Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg Pro Ala Leu Ser Gly
35 40 45

Asp Ala
50

<210> 699
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 699

Val Gln Lys Met Leu Gly His Cys Thr Phe Glu Pro Ala Glu Pro Lys
1 5 10 15

Ala Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys
20 25 30

Leu Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr
35 40 45

Asp Thr
50

<210> 700
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 700

Val Gln Lys Met Leu Gly His Cys Thr Phe Glu Pro Ala Glu Pro Lys
1 5 10 15

Ala Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys
20 25 30

Leu Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr
35 40 45

Asp Thr
50

<210> 701
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 701

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

Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe
20 25 30

Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met
35 40 45

Glu Met
50

<210> 702
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 702

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

Tyr Ala Gln Ala Arg Lys Leu Leu Ser Leu Glu Asp Thr Ala Phe Phe
20 25 30

Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met
35 40 45

Glu Met
50

<210> 703
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 703

Lys Ala Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys
1 5 10 15

Asp Lys Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile
20 25 30

Gly Thr Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg
35 40 45

Leu Lys
50

<210> 704
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 704

```
Lys Ala Tyr His Thr Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys
1               5                  10                  15

Asp Lys Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile
            20                  25                  30

Gly Thr Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg
        35                  40                  45

Leu Lys
    50
```

<210> 705
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 705

```
Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
1               5                  10                  15

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            20                  25                  30

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        35                  40                  45

Tyr Gly
    50
```

<210> 706
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 706

```
Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
1               5                  10                  15

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            20                  25                  30
```

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
35 40 45

Tyr Gly
50

<210> 707
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 707

Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro
1 5 10 15

Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser
20 25 30

Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro
35 40 45

Ala Arg
50

<210> 708
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 708

Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro
1 5 10 15

Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser
20 25 30

Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro
35 40 45

Ala Arg
50

<210> 709

<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 709

Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp Arg
1 5 10 15

Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu Lys
20 25 30

Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu Pro
35 40 45

Lys Ser
50

<210> 710
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 710

Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp Arg
1 5 10 15

Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu Lys
20 25 30

Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu Pro
35 40 45

Lys Ser
50

<210> 711
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 711

Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His Gly Lys Cys
1 5 10 15

Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly Arg Leu Asn Glu Lys Gly
20 25 30

Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp
35 40 45

Ser Phe
50

<210> 712
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 712

Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His Gly Lys Cys
1 5 10 15

Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly Arg Leu Asn Glu Lys Gly
20 25 30

Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp
35 40 45

Ser Phe
50

<210> 713
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 713

Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly Asn
1 5 10 15

Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu Trp
20 25 30

Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser Lys

35 40 45

Lys Gln
50

<210> 714
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 714

Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly Asn
1 5 10 15

Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu Trp
20 25 30

Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser Lys
35 40 45

Lys Gln
50

<210> 715
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 715

Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn
1 5 10 15

Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu Cys Gln Phe Val Ala
20 25 30

Asp Arg Met Arg Leu Thr Gly Lys Gly Lys Lys Arg Val Phe Ala Ser
35 40 45

Asn Gly
50

<210> 716
<211> 50

<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 716

Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn
1 5 10 15

Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu Cys Gln Phe Val Ala
20 25 30

Asp Arg Met Arg Leu Thr Gly Lys Gly Lys Lys Arg Val Phe Ala Ser
35 40 45

Asn Gly
50

<210> 717
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 717

Gln Ile Thr Asn Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg
1 5 10 15

Ala Glu Asn Asp Arg His His Ala Leu Asp Ala Val Val Val Ala Cys
20 25 30

Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys
35 40 45

Glu Met
50

<210> 718
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 718

Gln Ile Thr Asn Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg

1 5 10 15

Ala Glu Asn Asp Arg His His Ala Leu Asp Ala Val Val Val Ala Cys
20 25 30

Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys
35 40 45

Glu Met
50

<210> 719
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 719

Asn Ala Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu
1 5 10 15

His Gln Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu
20 25 30

Val Met Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu
35 40 45

Glu Ala
50

<210> 720
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 720

Asn Ala Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu
1 5 10 15

His Gln Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu
20 25 30

Val Met Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu
35 40 45

Glu Ala
    50

<210> 721
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 721

Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
1   5   10   15

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            20   25   30

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        35   40   45

Ser Ala
    50

<210> 722
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 722

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
1   5   10   15

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            20   25   30

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        35   40   45

Ser Ala
    50

<210> 723
<211> 50
<212> PRT

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 723

Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu Thr Gln
1 5 10 15

Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg Glu Pro
20 25 30

Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys Asp Asp
35 40 45

Pro Ala
50

<210> 724
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 724

Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu Thr Gln
1 5 10 15

Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg Glu Pro
20 25 30

Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys Asp Asp
35 40 45

Pro Ala
50

<210> 725
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 725

Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys Ala Gly Asn Arg
1 5 10 15

Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val Gln Lys Thr Gly
20 25 30

Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn Ala Thr Met Val
35 40 45

Arg Val
50

<210> 726
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 726

Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys Ala Gly Asn Arg
1 5 10 15

Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val Gln Lys Thr Gly
20 25 30

Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn Ala Thr Met Val
35 40 45

Arg Val
50

<210> 727
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 727

Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr Leu Val Pro Ile Tyr Ser
1 5 10 15

Trp Gln Val Ala Lys Gly Ile Leu Pro Asp Arg Ala Val Val Gln Gly
20 25 30

Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp Asp Ser Phe Asn Phe Lys
35 40 45

Phe Ser
    50

<210> 728
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 728

Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr Leu Val Pro Ile Tyr Ser
1   5   10   15

Trp Gln Val Ala Lys Gly Ile Leu Pro Asp Arg Ala Val Val Ala Tyr
            20   25   30

Ala Asp Glu Glu Asp Trp Thr Val Ile Asp Glu Ser Phe Arg Phe Lys
        35   40   45

Phe Val
    50

<210> 729
<211> 50
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 729

Leu His Pro Asn Asp Leu Val Glu Val Ile Thr Lys Lys Ala Arg Met
1   5   10   15

Phe Gly Tyr Phe Ala Ser Cys His Arg Gly Thr Gly Asn Ile Asn Ile
            20   25   30

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

Gly Ile
    50

<210> 730
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 730

Gly Val Lys Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu
1 5 10 15

Gly Lys Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
20 25 30

<210> 731
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 731

Gly Val Lys Thr Ala Leu Ser Phe Gln Lys Tyr Gln Ile Asp Glu Met
1 5 10 15

Gly Lys Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
20 25 30

<210> 732
<211> 1140
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 732

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65 70 75 80

```
Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
                85                  90                  95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
            100                 105                 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
        115                 120                 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
    130                 135                 140

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

Gly Val Ala Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
                165                 170                 175

Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
            180                 185                 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320
```

```
Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

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

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val
```

565 570 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
580 585 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
595 600 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625 630 635 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
645 650 655

Phe Asp Glu Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr
660 665 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr
675 680 685

Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
690 695 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705 710 715 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
725 730 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
740 745 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln
755 760 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
770 775 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785 790 795 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
805 810 815

```
Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Ala His Lys Asp Thr Leu Arg Ser
        835                 840                 845

Ala Lys Arg Phe Val Lys His Asn Glu Lys Ile Ser Val Lys Arg Val
    850                 855                 860

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

Lys Asn Gly Arg Glu Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg Leu
                885                 890                 895

Glu Ala Tyr Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn
            900                 905                 910

Pro Phe Tyr Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val Glu
        915                 920                 925

Lys Thr Gln Glu Ser Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr Thr
    930                 935                 940

Ile Ala Asp Asn Gly Asp Met Val Arg Val Asp Val Phe Cys Lys Val
945                 950                 955                 960

Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile Val Pro Ile Tyr Ala Trp
                965                 970                 975

Gln Val Ala Glu Asn Ile Leu Pro Asp Ile Asp Cys Lys Gly Tyr Arg
            980                 985                 990

Ile Asp Asp Ser Tyr Thr Phe Cys  Phe Ser Leu His Lys  Tyr Asp Leu
        995                 1000                1005

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

Tyr Ile  Asn Cys Asp Ser Ser  Asn Gly Arg Phe Tyr  Leu Ala Trp
    1025                1030                1035

His Asp  Lys Gly Ser Lys Glu  Gln Gln Phe Arg Ile  Ser Thr Gln
    1040                1045                1050
```

```
Asn Leu  Val Leu Ile Gln Lys  Tyr Gln Val Asn Glu  Leu Gly Lys
    1055                1060                1065

Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg Pro Pro  Val Arg Gly
    1070                1075                1080

Thr Gly  Gly Pro Lys Lys Lys  Arg Lys Val Tyr Pro  Tyr Asp Val
    1085                1090                1095

Pro Asp  Tyr Ala Gly Tyr Pro  Tyr Asp Val Pro Asp  Tyr Ala Gly
    1100                1105                1110

Ser Tyr  Pro Tyr Asp Val Pro  Asp Tyr Ala Gly Ser  Ala Ala Pro
    1115                1120                1125

Ala Ala  Lys Lys Lys Lys Leu  Asp Phe Glu Ser Gly
    1130                1135                1140
```

<210> 733
<211> 1168
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 733

```
Met Val Pro Lys Lys Lys Arg Lys Val Glu Asp Lys Arg Pro Ala Ala
1               5                   10                  15

Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Met Ala Ala Phe Lys
            20                  25                  30

Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Ala Ser
        35                  40                  45

Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu Glu Asn Pro Ile Arg
    50                  55                  60

Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg Ala Glu Val Pro Lys
65                  70                  75                  80

Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu Ala Arg Ser Val Arg
                85                  90                  95

Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala Arg Arg Leu
            100                 105                 110
```

```
Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp Phe Asp Glu Asn Gly
        115                 120                 125

Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala Ala Ala
    130                 135                 140

Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu His
145                 150                 155                 160

Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu Gly Glu
                165                 170                 175

Thr Ala Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Asn Asn
            180                 185                 190

Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala
        195                 200                 205

Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln Arg Gly
    210                 215                 220

Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu Ile
225                 230                 235                 240

Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser Gly
                245                 250                 255

Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg Pro Ala
            260                 265                 270

Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly His Cys Thr Phe Glu
        275                 280                 285

Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe
    290                 295                 300

Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser
305                 310                 315                 320

Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu Met Asp Glu Pro
                325                 330                 335

Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg Lys Leu Leu Gly
            340                 345                 350
```

```
Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr Gly Lys Asp Asn
        355                 360                 365

Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala Tyr His Ala Ile Ser
    370                 375                 380

Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn
385                 390                 395                 400

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

Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val Gln Pro
            420                 425                 430

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

Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val Pro Leu Met Glu Gln
    450                 455                 460

Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp His Tyr
465                 470                 475                 480

Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile Pro Ala
                485                 490                 495

Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser Gln Ala Arg
            500                 505                 510

Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro Ala Arg Ile
        515                 520                 525

His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp Arg Lys
    530                 535                 540

Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu Lys Ala
545                 550                 555                 560

Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu Pro Lys
                565                 570                 575

Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His Gly Lys
            580                 585                 590
```

```
Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg Leu Asn Glu Lys
        595                 600                 605

Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp
    610                 615                 620

Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn
625                 630                 635                 640

Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser
                645                 650                 655

Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro
            660                 665                 670

Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp Gly
        675                 680                 685

Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu
    690                 695                 700

Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr Gly Lys Gly Lys Arg
705                 710                 715                 720

Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn Leu Leu Arg Gly Phe
                725                 730                 735

Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp Arg His His Ala Leu
            740                 745                 750

Asp Ala Val Val Val Ala Cys Ser Thr Val Ala Met Gln Gln Lys Ile
        755                 760                 765

Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala Phe Asp Gly Lys Thr
    770                 775                 780

Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln Lys Thr His Phe Pro
785                 790                 795                 800

Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met Ile Arg Val Phe Gly
                805                 810                 815

Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala Asp Thr Pro Glu Lys
            820                 825                 830

Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser Arg Pro Glu Ala Val
```

```
        835                 840                 845

His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg Ala Pro Asn Arg Lys
    850                 855                 860

Met Ser Gly Ala His Lys Asp Thr Leu Arg Ser Ala Lys Arg Phe Val
865                 870                 875                 880

Lys His Asn Glu Lys Ile Ser Val Lys Arg Val Trp Leu Thr Glu Ile
                885                 890                 895

Lys Leu Ala Asp Leu Glu Asn Met Val Asn Tyr Lys Asn Gly Arg Glu
            900                 905                 910

Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala Tyr Gly Gly
        915                 920                 925

Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn Pro Phe Tyr Lys Lys
    930                 935                 940

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

Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr Thr Ile Ala Asp Asn Gly
                965                 970                 975

Asp Met Val Arg Val Asp Val Phe Cys Lys Val Asp Lys Lys Gly Lys
            980                 985                 990

Asn Gln Tyr Phe Ile Val Pro Ile  Tyr Ala Trp Gln Val  Ala Glu Asn
        995                 1000                1005

Ile Leu  Pro Asp Ile Asp Cys  Lys Gly Tyr Arg Ile  Asp Asp Ser
    1010                1015                1020

Tyr Thr  Phe Cys Phe Ser Leu  His Lys Tyr Asp Leu  Ile Ala Phe
    1025                1030                1035

Gln Lys  Asp Glu Lys Ser Lys  Val Glu Phe Ala Tyr  Tyr Ile Asn
    1040                1045                1050

Cys Asp  Ser Ser Asn Gly Arg  Phe Tyr Leu Ala Trp  His Asp Lys
    1055                1060                1065

Gly Ser  Lys Glu Gln Gln Phe  Arg Ile Ser Thr Gln  Asn Leu Val
    1070                1075                1080
```

Leu Ile Gln Lys Tyr Gln Val Asn Glu Leu Gly Lys Glu Ile Arg
1085 1090 1095

Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg Glu Asp Lys Arg
1100 1105 1110

Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Tyr
1115 1120 1125

Pro Tyr Asp Val Pro Asp Tyr Ala Gly Tyr Pro Tyr Asp Val Pro
1130 1135 1140

Asp Tyr Ala Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Ala
1145 1150 1155

Ala Pro Ala Ala Lys Lys Lys Lys Leu Asp
1160 1165

<210> 734
<211> 1135
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 734

Pro Lys Lys Lys Arg Lys Val Asn Ala Met Ala Ala Phe Lys Pro Asn
1 5 10 15

Pro Ile Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Ala Ser Val Gly
20 25 30

Trp Ala Met Val Glu Ile Asp Glu Glu Glu Asn Pro Ile Arg Leu Ile
35 40 45

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

Asp Ser Leu Ala Met Ala Arg Arg Leu Ala Arg Ser Val Arg Arg Leu
65 70 75 80

Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala Arg Arg Leu Leu Lys
85 90 95

Arg Glu Gly Val Leu Gln Ala Ala Asp Phe Asp Glu Asn Gly Leu Ile

100 105 110

Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala Ala Ala Leu Asp
115 120 125

Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu His Leu Ile
130 135 140

Lys His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu Gly Glu Thr Ala
145 150 155 160

Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Asn Asn Ala His
165 170 175

Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala Leu Asn
180 185 190

Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln Arg Gly Asp Tyr
195 200 205

Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu Ile Leu Leu
210 215 220

Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser Gly Gly Leu
225 230 235 240

Lys Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg Pro Ala Leu Ser
245 250 255

Gly Asp Ala Val Gln Lys Met Leu Gly His Cys Thr Phe Glu Pro Ala
260 265 270

Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe Ile Trp
275 280 285

Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser Glu Arg
290 295 300

Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu Met Asp Glu Pro Tyr Arg
305 310 315 320

Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu
325 330 335

Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu
340 345 350

```
Ala Ser Thr Leu Met Glu Met Lys Ala Tyr His Ala Ile Ser Arg Ala
        355                 360                 365

Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn Leu Ser
    370                 375                 380

Ser Glu Leu Gln Asp Glu Ile Gly Thr Ala Phe Ser Leu Phe Lys Thr
385                 390                 395                 400

Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val Gln Pro Glu Ile
                405                 410                 415

Leu Glu Ala Leu Leu Lys His Ile Ser Phe Asp Lys Phe Val Gln Ile
            420                 425                 430

Ser Leu Lys Ala Leu Arg Arg Ile Val Pro Leu Met Glu Gln Gly Lys
        435                 440                 445

Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp His Tyr Gly Lys
    450                 455                 460

Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile Pro Ala Asp Glu
465                 470                 475                 480

Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser Gln Ala Arg Lys Val
                485                 490                 495

Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro Ala Arg Ile His Ile
            500                 505                 510

Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp Arg Lys Glu Ile
        515                 520                 525

Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu Lys Ala Ala Ala
    530                 535                 540

Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu Pro Lys Ser Lys
545                 550                 555                 560

Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu
                565                 570                 575

Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg Leu Asn Glu Lys Gly Tyr
            580                 585                 590
```

```
Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser
        595                 600                 605

Phe Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly
    610                 615                 620

Asn Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu
625                 630                 635                 640

Trp Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser
                645                 650                 655

Lys Lys Gln Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp Gly Phe Lys
            660                 665                 670

Glu Cys Asn Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu Cys Gln
        675                 680                 685

Phe Val Ala Asp His Ile Leu Leu Thr Gly Lys Gly Lys Arg Arg Val
    690                 695                 700

Phe Ala Ser Asn Gly Gln Ile Thr Asn Leu Leu Arg Gly Phe Trp Gly
705                 710                 715                 720

Leu Arg Lys Val Arg Ala Glu Asn Asp Arg His His Ala Leu Asp Ala
                725                 730                 735

Val Val Val Ala Cys Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg
            740                 745                 750

Phe Val Arg Tyr Lys Glu Met Asn Ala Phe Asp Gly Lys Thr Ile Asp
        755                 760                 765

Lys Glu Thr Gly Lys Val Leu His Gln Lys Thr His Phe Pro Gln Pro
    770                 775                 780

Trp Glu Phe Phe Ala Gln Glu Val Met Ile Arg Val Phe Gly Lys Pro
785                 790                 795                 800

Asp Gly Lys Pro Glu Phe Glu Glu Ala Asp Thr Pro Glu Lys Leu Arg
                805                 810                 815

Thr Leu Leu Ala Glu Lys Leu Ser Ser Arg Pro Glu Ala Val His Glu
            820                 825                 830
```

```
Tyr Val Thr Pro Leu Phe Val Ser Arg Ala Pro Asn Arg Lys Met Ser
        835                 840                 845

Gly Ala His Lys Asp Thr Leu Arg Ser Ala Lys Arg Phe Val Lys His
    850                 855                 860

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

Ala Asp Leu Glu Asn Met Val Asn Tyr Lys Asn Gly Arg Glu Ile Glu
                885                 890                 895

Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala Tyr Gly Gly Asn Ala
            900                 905                 910

Lys Gln Ala Phe Asp Pro Lys Asp Asn Pro Phe Tyr Lys Lys Gly Gly
        915                 920                 925

Gln Leu Val Lys Ala Val Arg Val Glu Lys Thr Gln Glu Ser Gly Val
    930                 935                 940

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

Val Arg Val Asp Val Phe Cys Lys Val Asp Lys Lys Gly Lys Asn Gln
                965                 970                 975

Tyr Phe Ile Val Pro Ile Tyr Ala Trp Gln Val Ala Glu Asn Ile Leu
            980                 985                 990

Pro Asp Ile Asp Cys Lys Gly Tyr  Arg Ile Asp Asp Ser  Tyr Thr Phe
        995                 1000                1005

Cys Phe  Ser Leu His Lys Tyr  Asp Leu Ile Ala Phe  Gln Lys Asp
    1010                1015                1020

Glu Lys  Ser Lys Val Glu Phe  Ala Tyr Tyr Ile Asn  Cys Asp Ser
    1025                1030                1035

Ser Asn  Gly Arg Phe Tyr Leu  Ala Trp His Asp Lys  Gly Ser Lys
    1040                1045                1050

Glu Gln  Gln Phe Arg Ile Ser  Thr Gln Asn Leu Val  Leu Ile Gln
    1055                1060                1065

Lys Tyr  Gln Val Asn Glu Leu  Gly Lys Glu Ile Arg  Pro Cys Arg
```

1070 1075 1080

Leu Lys Lys Arg Pro Pro Val Arg Gly Gly Gly Gly Ser Gly Gly
1085 1090 1095

Gly Gly Ser Gly Gly Gly Gly Ser Pro Ala Ala Lys Lys Lys Lys
1100 1105 1110

Leu Asp Gly Gly Gly Ser Lys Arg Pro Ala Ala Thr Lys Lys Ala
1115 1120 1125

Gly Gln Ala Lys Lys Lys Lys
1130 1135

<210> 735
<211> 1135
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 735

Pro Lys Lys Lys Arg Lys Val Asn Ala Met Ala Ala Phe Lys Pro Asn
1 5 10 15

Pro Ile Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Ala Ser Val Gly
20 25 30

Trp Ala Met Val Glu Ile Asp Glu Glu Glu Asn Pro Ile Arg Leu Ile
35 40 45

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

Asp Ser Leu Ala Met Ala Arg Arg Leu Ala Arg Ser Val Arg Arg Leu
65 70 75 80

Thr Arg Arg Arg Ala His Arg Leu Leu Arg Ala Arg Arg Leu Leu Lys
85 90 95

Arg Glu Gly Val Leu Gln Ala Ala Asp Phe Asp Glu Asn Gly Leu Ile
100 105 110

Lys Ser Leu Pro Asn Thr Pro Trp Gln Leu Arg Ala Ala Ala Leu Asp
115 120 125

```
Arg Lys Leu Thr Pro Leu Glu Trp Ser Ala Val Leu Leu His Leu Ile
    130                 135                 140

Lys His Arg Gly Tyr Leu Ser Gln Arg Lys Asn Glu Gly Glu Thr Ala
145                 150                 155                 160

Asp Lys Glu Leu Gly Ala Leu Leu Lys Gly Val Ala Asn Asn Ala His
                165                 170                 175

Ala Leu Gln Thr Gly Asp Phe Arg Thr Pro Ala Glu Leu Ala Leu Asn
            180                 185                 190

Lys Phe Glu Lys Glu Ser Gly His Ile Arg Asn Gln Arg Gly Asp Tyr
        195                 200                 205

Ser His Thr Phe Ser Arg Lys Asp Leu Gln Ala Glu Leu Ile Leu Leu
    210                 215                 220

Phe Glu Lys Gln Lys Glu Phe Gly Asn Pro His Val Ser Gly Gly Leu
225                 230                 235                 240

Lys Glu Gly Ile Glu Thr Leu Leu Met Thr Gln Arg Pro Ala Leu Ser
                245                 250                 255

Gly Asp Ala Val Gln Lys Met Leu Gly His Cys Thr Phe Glu Pro Ala
            260                 265                 270

Glu Pro Lys Ala Ala Lys Asn Thr Tyr Thr Ala Glu Arg Phe Ile Trp
        275                 280                 285

Leu Thr Lys Leu Asn Asn Leu Arg Ile Leu Glu Gln Gly Ser Glu Arg
    290                 295                 300

Pro Leu Thr Asp Thr Glu Arg Ala Thr Leu Met Asp Glu Pro Tyr Arg
305                 310                 315                 320

Lys Ser Lys Leu Thr Tyr Ala Gln Ala Arg Lys Leu Leu Gly Leu Glu
                325                 330                 335

Asp Thr Ala Phe Phe Lys Gly Leu Arg Tyr Gly Lys Asp Asn Ala Glu
            340                 345                 350

Ala Ser Thr Leu Met Glu Met Lys Ala Tyr His Ala Ile Ser Arg Ala
        355                 360                 365

Leu Glu Lys Glu Gly Leu Lys Asp Lys Lys Ser Pro Leu Asn Leu Ser
```

370 375 380

Ser Glu Leu Gln Asp Glu Ile Gly Thr Ala Phe Ser Leu Phe Lys Thr
385 390 395 400

Asp Glu Asp Ile Thr Gly Arg Leu Lys Asp Arg Val Gln Pro Glu Ile
405 410 415

Leu Glu Ala Leu Leu Lys His Ile Ser Phe Asp Lys Phe Val Gln Ile
420 425 430

Ser Leu Lys Ala Leu Arg Arg Ile Val Pro Leu Met Glu Gln Gly Lys
435 440 445

Arg Tyr Asp Glu Ala Cys Ala Glu Ile Tyr Gly Asp His Tyr Gly Lys
450 455 460

Lys Asn Thr Glu Glu Lys Ile Tyr Leu Pro Pro Ile Pro Ala Asp Glu
465 470 475 480

Ile Arg Asn Pro Val Val Leu Arg Ala Leu Ser Gln Ala Arg Lys Val
485 490 495

Ile Asn Gly Val Val Arg Arg Tyr Gly Ser Pro Ala Arg Ile His Ile
500 505 510

Glu Thr Ala Arg Glu Val Gly Lys Ser Phe Lys Asp Arg Lys Glu Ile
515 520 525

Glu Lys Arg Gln Glu Glu Asn Arg Lys Asp Arg Glu Lys Ala Ala Ala
530 535 540

Lys Phe Arg Glu Tyr Phe Pro Asn Phe Val Gly Glu Pro Lys Ser Lys
545 550 555 560

Asp Ile Leu Lys Leu Arg Leu Tyr Glu Gln Gln His Gly Lys Cys Leu
565 570 575

Tyr Ser Gly Lys Glu Ile Asn Leu Val Arg Leu Asn Glu Lys Gly Tyr
580 585 590

Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser
595 600 605

Phe Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys Gly
610 615 620

```
Asn Gln Thr Pro Tyr Glu Tyr Phe Asn Gly Lys Asp Asn Ser Arg Glu
625                 630                 635                 640

Trp Gln Glu Phe Lys Ala Arg Val Glu Thr Ser Arg Phe Pro Arg Ser
                645                 650                 655

Lys Lys Gln Arg Ile Leu Leu Gln Lys Phe Asp Glu Asp Gly Phe Lys
            660                 665                 670

Glu Cys Asn Leu Asn Asp Thr Arg Tyr Val Asn Arg Phe Leu Cys Gln
        675                 680                 685

Phe Val Ala Asp His Ile Leu Leu Thr Gly Lys Gly Lys Arg Arg Val
    690                 695                 700

Phe Ala Ser Asn Gly Gln Ile Thr Asn Leu Leu Arg Gly Phe Trp Gly
705                 710                 715                 720

Leu Arg Lys Val Arg Ala Glu Asn Asp Arg His His Ala Leu Asp Ala
                725                 730                 735

Val Val Val Ala Cys Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg
            740                 745                 750

Phe Val Arg Tyr Lys Glu Met Asn Ala Phe Asp Gly Lys Thr Ile Asp
        755                 760                 765

Lys Glu Thr Gly Lys Val Leu His Gln Lys Thr His Phe Pro Gln Pro
    770                 775                 780

Trp Glu Phe Phe Ala Gln Glu Val Met Ile Arg Val Phe Gly Lys Pro
785                 790                 795                 800

Asp Gly Lys Pro Glu Phe Glu Glu Ala Asp Thr Pro Glu Lys Leu Arg
                805                 810                 815

Thr Leu Leu Ala Glu Lys Leu Ser Ser Arg Pro Glu Ala Val His Glu
            820                 825                 830

Tyr Val Thr Pro Leu Phe Val Ser Arg Ala Pro Asn Arg Lys Met Ser
        835                 840                 845

Gly Ala His Lys Asp Thr Leu Arg Ser Ala Lys Arg Phe Val Lys His
    850                 855                 860
```

```
Asn Glu Lys Ile Ser Val Lys Arg Val Trp Leu Thr Glu Ile Lys Leu
865                 870                 875                 880

Ala Asp Leu Glu Asn Met Val Asn Tyr Lys Asn Gly Arg Glu Ile Glu
                885                 890                 895

Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala Tyr Gly Gly Asn Ala
            900                 905                 910

Lys Gln Ala Phe Asp Pro Lys Asp Asn Pro Phe Tyr Lys Lys Gly Gly
        915                 920                 925

Gln Leu Val Lys Ala Val Arg Val Glu Lys Thr Gln Glu Ser Gly Val
    930                 935                 940

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

Val Arg Val Asp Val Phe Cys Lys Val Asp Lys Lys Gly Lys Asn Gln
                965                 970                 975

Tyr Phe Ile Val Pro Ile Tyr Ala Trp Gln Val Ala Glu Asn Ile Leu
            980                 985                 990

Pro Asp Ile Asp Cys Lys Gly Tyr  Arg Ile Asp Asp Ser  Tyr Thr Phe
        995                 1000                1005

Cys Phe  Ser Leu His Lys Tyr  Asp Leu Ile Ala Phe  Gln Lys Asp
    1010                1015                1020

Glu Lys  Ser Lys Val Glu Phe  Ala Tyr Tyr Ile Asn  Cys Asp Ser
    1025                1030                1035

Ser Asn  Gly Arg Phe Tyr Leu  Ala Trp His Asp Lys  Gly Ser Lys
    1040                1045                1050

Glu Gln  Gln Phe Arg Ile Ser  Thr Gln Asn Leu Val  Leu Ile Gln
    1055                1060                1065

Lys Tyr  Gln Val Asn Glu Leu  Gly Lys Glu Ile Arg  Pro Cys Arg
    1070                1075                1080

Leu Lys  Lys Arg Pro Pro Val  Arg Gly Gly Gly Gly  Ser Gly Gly
    1085                1090                1095
```

Gly Gly Ser Gly Gly Gly Gly Ser Pro Ala Ala Lys Lys Lys Lys
1100 1105 1110

Leu Asp Gly Gly Gly Ser Lys Arg Pro Ala Ala Thr Lys Lys Ala
1115 1120 1125

Gly Gln Ala Lys Lys Lys Lys
1130 1135

<210> 736
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 736
gtcacctgcc tcgtggaata cgg 23

<210> 737
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 737
gcacctgcct cgtggaatac ggt 23

<210> 738
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 738
gttcagcgtg tccggctttg gc 22

<210> 739
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 739
gtggtgagca agggcgagga gctg 24

<210> 740

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 740
gggcgaggag ctgttcaccg gggt 24

<210> 741
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 741
gggcgaggag ctgttcaccg gggt 24

<210> 742
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 742
gcgtccagct cgaccaggat gggc 24

<210> 743
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 743
gcggtgaaca gctcctcgcc cttg 24

<210> 744
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 744
gggcaccacc ccggtgaaca gctc 24

<210> 745
<211> 24

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 745
ggcaccaccc cggtgaacag ctcc 24

<210> 746
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 746
gggatgggca ccaccccggt gaac 24

<210> 747
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 747
gcgtgtccgg ctttggcgag acaa 24

<210> 748
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 748
gtccggcttt ggcgagacaa atca 24

<210> 749
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 749
gatcacctgc ctcgtggaat acgg 24

<210> 750
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 750
gatcacctgc ctcgtggaat acgg 24

<210> 751
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 751
gccaaagccg gacacgctga actt 24

<210> 752
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 752
ggaacttgtg gccgtttacg tcg 23

<210> 753
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 753
gaacttgtgg ccgtttacgt cg 22

<210> 754
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 754
gacttgtggc cgtttacgtc g 21

<210> 755
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 755
gcttgtggcc gtttacgtcg 20

<210> 756
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 756
gttgtggccg tttacgtcg 19

<210> 757
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 757
gtcacctgcc tcgtggaata cgg 23

<210> 758
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 758
gcacctgcct cgtggaatac gg 22

<210> 759
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 759
gacctgcctc gtggaatacg g 21

<210> 760
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 760
gcctgcctcg tggaatacgg 20

<210> 761
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 761
ggttctgggt acttttatct gtcc 24

<210> 762
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 762
gtctgcctaa caggaggtgg gggt 24

<210> 763
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 763
gaatatcagg agactaggaa ggag 24

<210> 764
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 764
gcctccctgc agggctgctc cc 22

<210> 765
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 765
gagctagtct tcttcctcca accc 24

<210> 766
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 766
gatctgtccc ctccacccca cagt 24

<210> 767
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 767
gatctgtccc ctccacccca cagt 24

<210> 768
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 768
gcatcctctt gctttctttg cctg 24

<210> 769
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 769
ggagtcgcca gaggccggtg gtgg 24

<210> 770
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 770
gcccagcggc cggatatcag ctgc 24

<210> 771
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 771
ggaagggaac atattactat tgc 23

<210> 772
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 772
gtggagtggc ctgctatcag ctac 24

<210> 773
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 773
gaggaaggga acatattact attg 24

<210> 774
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 774
gtgaattctc atcagctaaa atgc 24

<210> 775
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 775
gctcactcac ccacacagac acac 24

<210> 776
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 776
ggaagaattt cattctgttc tcag 24

<210> 777
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 777
gctcagtttt cctggattat gcct 24

<210> 778
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 778
gcgttggagc ggggagaagg ccag 24

<210> 779
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 779
gggccgcgga gatagctgca gggc 24

<210> 780
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 780

gcccacccgg cggcgcctcc ctgc 24

<210> 781
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 781
gcgtggcagc tgatatccgg ccgc 24

<210> 782
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 782
gccgcggcgc gacgtggagc cagc 24

<210> 783
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 783
gtgctcccca gcccaaaccg ccgc 24

<210> 784
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 784
gtcagattgg cttgctcgga attg 24

<210> 785
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 785
gctgggtgaa tggagcgagc agcg 24

<210> 786
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 786
gtcctggagt gacccctggc cttc 24

<210> 787
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 787
gatcctggag tgacccctgg cctt 24

<210> 788
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 788
gtgtgtccct ctccccaccc gtcc 24

<210> 789
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 789
gttggagcgg ggagaaggcc aggg 24

<210> 790
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 790
gcgttggagc ggggagaagg ccag 24

<210> 791
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 791
gtaccctcca ataatttggc tggc 24

<210> 792
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 792
gataatttgg ctggcaattc cgag 24

<210> 793
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 793
gaaaattgtg atttccagat ccac 24

<210> 794
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 794
gagcagaaaa aattgtgatt tcc 23

<210> 795
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 795
gcaggggcca ggtgtccttc tctg 24

<210> 796
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 796
gaatggcagg cggaggttgt actg 24

<210> 797
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 797
gagtgagaga gtgagagaga gaca 24

<210> 798
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 798
gtgagcaggc acctgtgcca acat 24

<210> 799
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 799
gcgtgggggc tccgtgcccc acgc 24

<210> 800
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 800
gcatgggcag gggctggggt gcac 24

<210> 801

<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 801
gggccaggtg tccttctctg 20

<210> 802
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 802
ggcaggcgga ggttgtactg 20

<210> 803
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 803
gagagagtga gagagagaca 20

<210> 804
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 804
gcaggcacct gtgccaacat 20

<210> 805
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 805
gggggctccg tgccccacgc 20

<210> 806
<211> 20

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 806
gggcaggggc tggggtgcac 20

<210> 807
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 807
tggcttagca cctctccat 19

<210> 808
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 808
agaggagcct tctgactgct gcaga 25

<210> 809
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 809
tccgtcttcc tccactcc 18

<210> 810
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 810
tagagaactg ggtagtgtg 19

<210> 811
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 811
gtacatgaag caactccagt ccca 24

<210> 812
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 812
tggtgattat gggagaactg gagc 24

<210> 813
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 813
ggcataagga aatcgaaggt c 21

<210> 814
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 814
acacgggcag catgggaata gtc 23

<210> 815
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 815
cctgtgtggc tttgctttgg tcg 23

<210> 816
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 816
ggaggaagag tagctcgccg agg 23

<210> 817
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 817
agggagaggg aagtgtgggg aagg 24

<210> 818
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 818
agaactcagg accaacttat tctg 24

<210> 819
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 819
atgacagaca caaccagagg gca 23

<210> 820
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 820
taggaaggag gaggcctaag 20

<210> 821
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 821
ccaatattgc atgggatgg 19

<210> 822
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 822
atcaaattcc agcaccgagc gc 22

<210> 823
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 823
accattgagg acgtttgtct cac 23

<210> 824
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 824
catgtcctca agtcaagaac aag 23

<210> 825
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 825
gctaggggag agtcccactg tcca 24

<210> 826
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 826
gtagggtgtg atgggaggct aagc 24

<210> 827
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 827
agaccgagtg gcagtgacag caag 24

<210> 828
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 828
gtcttcctgc tctgtgcgca cgac 24

<210> 829
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (35)..(44)
<223> n is a, c, g, or t

<400> 829
tagcggccgc tcatgcgcgg cgcattacct ttacnnnnnn nnnnggatcc tctagagtcg 60

<210> 830
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 830
acaggaaaca gctatgacca tgaaagcttg catgcctgca ggtcgactct agaggatc 58

<210> 831

<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 831
ctacacgacg ctcttccgat ctcctggagc gtgtacgttg g 41

<210> 832
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 832
ctacacgacg ctcttccgat ctcctgtggt cccagctact tg 42

<210> 833
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 833
ctacacgacg ctcttccgat ctatctgcga tgtcctcgag g 41

<210> 834
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 834
ctacacgacg ctcttccgat cttggtgtgc gcctctaacg 40

<210> 835
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 835
ctacacgacg ctcttccgat ctggagtctt gctttgtcac tcaga 45

<210> 836
<211> 41

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 836
ctacacgacg ctcttccgat ctagcctaga cccagtccca t 41

<210> 837
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 837
ctacacgacg ctcttccgat ctgctgggca tagtagtgga ct 42

<210> 838
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 838
ctacacgacg ctcttccgat cttggggagg ctgagacacg a 41

<210> 839
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 839
ctacacgacg ctcttccgat ctcttgggag gctgaggcaa g 41

<210> 840
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 840
agacgtgtgc tcttccgatc tcaggaggat gagagccagg 40

<210> 841
<211> 43
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 841
agacgtgtgc tcttccgatc tcagggtctc actctatcac cca 43

<210> 842
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 842
agacgtgtgc tcttccgatc tactgaatgg gttgaacttg gc 42

<210> 843
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 843
agacgtgtgc tcttccgatc tgagacagaa tcttgctctg tctcc 45

<210> 844
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 844
agacgtgtgc tcttccgatc ttcccagcta cttgggaggc 40

<210> 845
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 845
agacgtgtgc tcttccgatc tcctgcccaa atagggaagc ag 42

<210> 846
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 846
agacgtgtgc tcttccgatc ttggcgcctt agtctctgct ac 42

<210> 847
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 847
agacgtgtgc tcttccgatc tgcatgagac acagtttcac tctg 44

<210> 848
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 848
agacgtgtgc tcttccgatc tgagagagtc tcactgcgtt gc 42

<210> 849
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 849
ctacacgacg ctcttccgat cttctctcac ccactgggca c 41

<210> 850
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 850
agacgtgtgc tcttccgatc tgcttccaga cgagtgcaga 40

<210> 851
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 851
ctacacgacg ctcttccgat ctaagttttc aaaccagaag aactacgac 49

<210> 852
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 852
ctacacgacg ctcttccgat ctccggtata agtcctggag cg 42

<210> 853
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 853
ctacacgacg ctcttccgat ctccggtata agtcctggag cg 42

<210> 854
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 854
ctacacgacg ctcttccgat ctcctcgaat tccacggggt t 41

<210> 855
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 855
ctacacgacg ctcttccgat ctgttggtgg gagggaagtg ag 42

<210> 856
<211> 40
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 856
ctacacgacg ctcttccgat ctgatggcgg ttgtagcggc 40

<210> 857
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 857
ctacacgacg ctcttccgat ctcacataaa cctatgtttc agcaga 46

<210> 858
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 858
ctacacgacg ctcttccgat ctgctagttg gattgaagca gggt 44

<210> 859
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 859
ctacacgacg ctcttccgat ctttgagtgc ggcagcttcc 40

<210> 860
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 860
ctacacgacg ctcttccgat ctataaccct cccaggcaaa gtc 43

<210> 861
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 861
ctacacgacg ctcttccgat ctagcctgca catctgagct c 41

<210> 862
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 862
ctacacgacg ctcttccgat ctggagcatt gaagtgcctg g 41

<210> 863
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 863
ctacacgacg ctcttccgat ctggagcatt gaagtgcctg g 41

<210> 864
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 864
agacgtgtgc tcttccgatc tcatcctcga cagtcgcgg 39

<210> 865
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 865
agacgtgtgc tcttccgatc tgactgatca agtagaatac tcatggg 47

<210> 866
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 866
agacgtgtgc tcttccgatc tccctgccag cactgaagc 39

<210> 867
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 867
agacgtgtgc tcttccgatc tggttcctat ctttctagac caggagt 47

<210> 868
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 868
agacgtgtgc tcttccgatc tagtgtggag ggctcaggg 39

<210> 869
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 869
agacgtgtgc tcttccgatc tgatgggcag aggaaggcaa 40

<210> 870
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 870
agacgtgtgc tcttccgatc ttcactctca tgagcgtccc a 41

<210> 871
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 871

ctacacgacg ctcttccgat ctaaggttcc ttgcggttcg c 41

<210> 872
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 872
agacgtgtgc tcttccgatc tcgctgccat tgctccct 38

<210> 873
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 873
ctacacgacg ctcttccgat cttctcgcac attcttcacg tcc 43

<210> 874
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 874
agacgtgtgc tcttccgatc taggaacctt cccgacttag gg 42

<210> 875
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 875
ctacacgacg ctcttccgat ctcccgccca tcttctagaa agac 44

<210> 876
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 876
ctacacgacg ctcttccgat cttgccaggt gagggactgg 40

<210> 877
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 877
agacgtgtgc tcttccgatc ttctgggagt tctctgctgc c 41

<210> 878
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 878
agacgtgtgc tcttccgatc ttgcccaacc ttagcaagga g 41

<210> 879
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 879
ctacacgacg ctcttccgat cttaccttgg agcaacggcg 40

<210> 880
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 880
agacgtgtgc tcttccgatc tcccaggacg aggatggag 39

<210> 881
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 881
gatagtcact ccaggggttg 20

<210> 882
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 882
gtggtgaacc aatcagtcct 20

<210> 883
<211> 60
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 883

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met
50 55 60

<210> 884
<211> 49
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 884

Leu Tyr Ser Asn Asp Leu Ile Lys Val Gln Leu Lys Lys Asp Ser Phe
1 5 10 15

Leu Gly Tyr Phe Ser Gly Leu Asp Arg Ala Thr Gly Ala Ile Ser Leu
20 25 30

Arg Glu His Asp Leu Glu Lys Ser Lys Gly Lys Asp Gly Met His Arg
35 40 45

Ile

<210> 885
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 885
tccagctcga ccaggatggg caccaccccg 30

<210> 886
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 886
tccagctcga ccaggatggg caccaccccg 30

<210> 887
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 887
ttcagctcga ccaggatggg caccaccccg 30

<210> 888
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 888
tccagcttga ccaggatggg caccaccccg 30

<210> 889
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 889

ttcagcttga ccaggatggg caccaccccg 30

<210> 890
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 890
ggtcgagctg gacggcgacg taaacggcca 30

<210> 891
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 891
ggtcgagctg gacggcgacg taaacggcca 30

<210> 892
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 892
ggttgagctg gacggcgacg taaacggcca 30

<210> 893
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 893
agttcatctg caccaccggc aagctgcccg 30

<210> 894
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 894
agttcatctg caccaccggc aagctgcccg 30

<210> 895
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 895
agtttatctg caccaccggc aagctgcccg 30

<210> 896
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 896
agttcatttg caccaccggc aagctgcccg 30

<210> 897
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 897
cgtgccctgg cccaccctcg tgaccaccct 30

<210> 898
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 898
cgtgccctgg cccaccctcg tgaccaccct 30

<210> 899
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 899
cgtgccttgg cccaccctcg tgaccaccct 30

<210> 900
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 900
ggcctttccc tgtagccctg gggggagcca 30

<210> 901
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 901
ggcctttccc tgtagccctg gggggagcca 30

<210> 902
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 902
ggcctttccc tgtagccctg gggggagcca 30

<210> 903
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 903
ggcctttttc tgtagccctg gggggagcca 30

<210> 904
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 904
ggcctttacc tgtagccctg gggggagcca 30

<210> 905
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 905
ggccttttttt tgtagccctg gggggagcca 30

<210> 906
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 906
ggcctttgcc tgtagccctg gggggagcca 30

<210> 907
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 907
ggccttttct tgtagccctg gggggagcca 30

<210> 908
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 908
ggcctttcct gtagccctgg ggggagcca 29

<210> 909
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 909
ggcctttcct tgtagccctg gggggagcca 30

<210> 910

<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 910
ggcctttctc tgtagccctg gggggagcca 30

<210> 911
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 911
cagggctaca gggaaaggcc gaggaaacct 30

<210> 912
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 912
cagggctaca gggaaaggcc gaggaaacct 30

<210> 913
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 913
cagggctgca gggaaaggcc gaggaaacct 30

<210> 914
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 914
cagggctaca gggaaaggcc gaggaaacct 30

<210> 915
<211> 30

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 915
cagggctaca gggaaaggcc gaggaaacct 30

<210> 916
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 916
cagggctgca gggaaaggcc gaggaaacct 30

<210> 917
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 917
aacgactgga gcagtaatgc ctggtggccc agcttcctct 40

<210> 918
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 918
aacgactgga gcggtaatgc ctggtggccc agcttcctct 40

<210> 919
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 919
gtgaacttgt ggccgtttac gtcg 24

<210> 920
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 920
gtgaacttgt ggccgtttac gtcc 24

<210> 921
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 921
gtgaacttgt ggccgtttac gtgg 24

<210> 922
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 922
gtgaacttgt ggccgtttac gacg 24

<210> 923
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 923
gtgaacttgt ggccgtttac ctcg 24

<210> 924
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 924
gtgaacttgt ggccgtttag gtcg 24

<210> 925
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 925
gtgaacttgt ggccgttttc gtcg 24

<210> 926
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 926
gtgaacttgt ggccgttaac gtcg 24

<210> 927
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 927
gtgaacttgt ggccgtatac gtcg 24

<210> 928
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 928
gtgaacttgt ggccgattac gtcg 24

<210> 929
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 929
gtgaacttgt ggcccttac gtcg 24

<210> 930
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 930
gtgaacttgt ggcggtttac gtcg 24

<210> 931
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 931
gtgaacttgt gggcgtttac gtcg 24

<210> 932
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 932
gtgaacttgt gcccgtttac gtcg 24

<210> 933
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 933
gtgaacttgt cgccgtttac gtcg 24

<210> 934
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 934
gtgaacttga ggccgtttac gtcg 24

<210> 935
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 935
gtgaacttct ggccgtttac gtcg 24

<210> 936
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 936
gtgaactagt ggccgtttac gtcg 24

<210> 937
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 937
gtgaacatgt ggccgtttac gtcg 24

<210> 938
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 938
gtgaagttgt ggccgtttac gtcg 24

<210> 939
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 939
gtgatcttgt ggccgtttac gtcg 24

<210> 940
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 940
gtgtacttgt ggccgtttac gtcg 24

<210> 941
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 941
gtcaacttgt ggccgtttac gtcg 24

<210> 942
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 942
gagaacttgt ggccgtttac gtcg 24

<210> 943
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 943
ctcactcacc cacacagaca cacacgtcct 30

<210> 944
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 944
cacacacacc cacacagaca cacccccccc 30

<210> 945
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 945
ctcccctcaca cacacagaca cacacctccc 30

<210> 946
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 946
cacacacaca cacacagaca cacacacccc 30

<210> 947
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 947
gtgtgtccct ctccccaccc gtccctgtcc g 31

<210> 948
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 948
ttgtctccct gtccccaccc gtccccttca g 31

<210> 949
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 949
ctgcctccct ctgcccaccc gtccttccca c 31

<210> 950
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 950

ctgtgcctct ctccccaccc ttccacaccc t 31

<210> 951
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 951
gctcatcccc ctccccaccc gtcctcgccc g 31

<210> 952

<400> 952
000

<210> 953
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 953
tgaggaccgc cctgggcctg ggagaatccc t 31

<210> 954
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 954
gaaggaccac cctaggcctg ggagactccc t 31

<210> 955
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 955
tcctatggtg tctagaatgg tgagcaaggg cgaggaggac aacatggcca tcatcaagga 60

gttcatgcgc ttcaaggtgc acatggaggg ctccgtgaac 100

<210> 956
<211> 100
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 956
aggataccac agatcttacc actcgttccc gctcctcctg ttgtaccggt agtagttcct 60

caagtacgcg aagttccacg tgtacctccc gaggcacttg 100

<210> 957
<211> 33
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 957

Pro Met Val Ser Arg Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala
1 5 10 15

Ile Ile Lys Glu Phe Met Arg Phe Lys Val His Met Glu Gly Ser Val
20 25 30

Asn

<210> 958
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 958
ggccacgagt tcgagatcga gggtgagggt gagggccgac cctacgaggg cacctagacc 60

gccaagctga aggtgaccaa gggcggaccc ctgcccttcg 100

<210> 959
<211> 98
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 959
ccggtgctca agctctagct cccactccca ctcccggctg ggatgctccc gtggatctgg 60

cggttcgact tccactggtt cccgcctggg acggaagc 98

<210> 960
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 960

Gly His Glu Phe Glu Ile Glu Gly Glu Gly Glu Gly Arg Pro Tyr Glu
1 5 10 15

Gly Thr Thr Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe
20 25 30

<210> 961
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 961
cccgagcagg gccgcatgac caacaagatg aagagcacca aaggcgccct gaccttcagc 60

ccctacctgc tgagccacgt gatgggccac ggcttctacc 100

<210> 962
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 962
gggctcgtcc cggcgtactg gttgttctac ttctcgtggt ttccgcggga ctggaagtcg 60

gggatggacg actcggtgca ctacccggtg ccgaagatgg 100

<210> 963
<211> 33
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 963

Pro Glu Gln Gly Arg Met Thr Asn Lys Met Lys Ser Thr Lys Gly Ala
1 5 10 15

Leu Thr Phe Ser Pro Tyr Leu Leu Ser His Val Met Gly His Gly Phe

20 25 30

Tyr

<210> 964
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 964
acttcggcac ctaccccagc gggtacgaga acccccttcct gcacgccatc aacaacggcg 60

gctacaccaa cacccgcatc gagaagtacg aggacggcgg 100

<210> 965
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 965
tgaagccgtg gatggggtcg cccatgctct tggggaagga cgtgcggtag ttgttgccgc 60

cgatgtggtt gtgggcgtag ctcttcatgc tcctgccgcc 100

<210> 966
<211> 34
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 966

His Phe Gly Thr Tyr Pro Ser Gly Tyr Glu Asn Pro Phe Leu His Ala
1 5 10 15

Ile Asn Asn Gly Gly Tyr Thr Asn Thr Arg Ile Glu Lys Tyr Glu Asp
20 25 30

Gly Gly

<210> 967
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 967
gactcttctg catgggtgat gtcaatgcc 29

<210> 968
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 968
ccgccagaac cgagccggct actcggccc 29

<210> 969
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 969
cacacactca ctcacccaca cagacacaca cgtcctcac 39

<210> 970
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 970
gtgtgtgagt gagtgggtgt gtctgtgtgt gcaggagtg 39

<210> 971
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 971
gcucacucac ccacacagac acacguugu 29

CLAIMS

1. A mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for an N_4CC nucleotide sequence.

2. A protein according to Claim 1, in which the protein is Nme2Cas9.

3. A protein according to Claim 1, further comprising a nuclear localization signal protein.

4. A protein according to Claim 1, in which the nucleotide deaminase is a cytidine deaminase.

5. A protein according to Claim 1, in which the nucleotide deaminase is an adenosine deaminase.

6. A protein according to Claim 1, further comprising a uracil glycosylase inhibitor.

7. A protein according to Claim 1, in which the nuclear localization signal protein is selected from a nucleoplasmin and an SV40 .

8. A protein according to Claim 1, in which the binding region is a protospacer accessory motif interacting domain.

9. A protein according to Claim 8, in which the protospacer accessory motif interacting domain comprises the mutation.

10. A protein according to Claim 9, in which the mutation is a D16A mutation.

11. An adeno-associated virus comprising a mutated NmeCas9 protein, the mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for an N_4CC nucleotide sequence.

12. A virus according to Claim 11, in which the virus is an adeno-associated virus 8.

13. A virus according to Claim 11, in which the virus is an adeno-associated virus 6.

14. A virus according to Claim 11, in which the protein is Nme2Cas9.

15. A virus according to Claim 11, in which the protein further comprising a nuclear localization signal protein.

16. A virus according to Claim 11, in which the nucleotide deaminase is a cytidine deaminase.

17. A virus according to Claim 11, in which the nucleotide deaminase is an adenosine deaminase.

18. A virus according to Claim 11, in which the protein further comprises a uracil glycosylase inhibitor.

19. A virus according to Claim 11, in which the nuclear localization signal protein is selected from a nucleoplasmin and SV40.

20. A virus according to Claim 11, in which the binding region is a protospacer accessory motif interacting domain.

21. A virus according to Claim 20, in which the protospacer accessory motif interacting domain comprises the mutation.

22. A virus according to Claim 21, in which the mutation is a D16A mutation.

23. A method, comprising:

(a) providing;

(i) a nucleotide sequence comprising a gene with a mutated single base, in which the gene is flanked by an N_4CC nucleotide sequence;

(ii) a mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for the N_4CC nucleotide sequence;

(b) contacting the nucleotide sequence with the mutated NmeCas9 protein under conditions such that the binding region attaches to the N_4CC nucleotide sequence; and

(c) replacing the mutated single base with a wild type base with the mutated NmeCas9 protein.

24. A method according to Claim 23, in which the protein is Nme2Cas9.

25. A method according to Claim 23, in which the protein further comprising a nuclear localization signal protein.

26. A method according to Claim 23, in which the nucleotide deaminase is a cytidine deaminase.

27. A method according to Claim 23, in which the nucleotide deaminase is an adenosine deaminase.

28. A method according to Claim 23, in which the protein further comprises a uracil glycosylase inhibitor.

29. A method according to Claim 23, in which the nuclear localization signal protein is selected from the group consisting of nucleoplasmin and SV40.

30. A method according to Claim 23, in which the binding region is a protospacer accessory motif interacting domain.

31. A method according to Claim 30, in which the protospacer accessory motif interacting domain comprises the Cas9 protein mutation.

32. A method according to Claim 31, in which the Cas9 protein mutation is a D16A mutation.

33. A method, comprising:

(a) providing;

(i) a patient comprising a nucleotide sequence comprising a gene with a mutated single base, in which the gene is flanked by an N_4CC nucleotide sequence, in which the mutated gene causes a genetically-based medical condition;

(ii) an adeno-associated virus comprising a mutated NmeCas9 protein, the mutated NmeCas9 protein comprising a fused nucleotide deaminase and a binding region for the N_4CC nucleotide sequence;

(b) treating the patient with the adeno-associated virus under conditions such that the mutated NmeCas9 protein replaces the mutated single base with a wild type single base, such that the genetically-based medical condition does not develop.

34. A method according to Claim 33, in which the gene encodes a tyrosinase protein.

35. A method according to Claim 33, in which the genetically-based medical condition is tyrosinemia.

36. A method according to Claim 33, in which the virus is an adeno-associated virus 8.

37. A method according to Claim 33, in which the virus is an adeno-associated virus 6.

38. A method according to Claim 33, in which the protein is Nme2Cas9.

39. A method according to Claim 33, in which the protein further comprises a nuclear localization signal protein.

40. A method according to Claim 33, in which the nucleotide deaminase is a cytidine deaminase.

41. A method according to Claim 33, in which the nucleotide deaminase is an adenosine deaminase.

42. A method according to Claim 33, in which the protein further comprises a uracil glycosylase inhibitor.

43. A method according to Claim 33, in which the nuclear localization signal protein is selected from the group consisting of nucleoplasmin and SV40.

44. A method according to Claim 33, in which the binding region is a protospacer accessory motif interacting domain.

45. A method according to Claim 44, in which the protospacer accessory motif interacting domain comprises the mutation.

46. A method according to Claim 45, in which the mutation is a D16A mutation.