Targeted CRISPR Delivery Platforms

Field Of The Invention

The present invention is related to compositions and methods for gene therapy. Several approaches described herein utilize the *Neisseria meningitidis* Cas9 systems that provide hyperaccurate CRISPR gene editing platforms. Furthermore, the invention incorporates improvements of this Cas9 system: for example, truncating the single guide RNA sequences, and the packing of Nme1Cas9 or Nme2Ca9 with its guide RNA in an adeno-associated viral vector that is compatible for *in vivo* administration. Furthermore, Type II-C Cas9 orthologs have been identified that target protospacer adjacent motif sequences limited to between one – four required nucleotides.

Background

Clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR associated (Cas) is a unique RNA-guided adaptive immune system found in archaea and bacteria. These systems provide immunity by targeting and inactivating nucleic acids that originate from foreign genetic elements. Many different types of CRISPR-Cas systems have been identified to date and are categorized into two classes.

Within class II CRISPR systems, type II CRISPR-Cas systems are characterized by a single effector protein called Cas9, which forms a ribonucleoprotein (RNP) complex with CRISPR RNA (crRNA) and trans-activating RNA (tracrRNA) to target and cleave DNA. The crRNA contains a programmable guide sequence that can direct Cas9 to almost any DNA sequence in living organisms.

This programmability of Cas9 RNP complexes has been harnessed by many researchers for genome editing in eukaryotic systems. It has been used to edit the genomes of mammalian cells, human embryos, plants, rodents, and other living organisms. Cas9 RNPs have been used for precise (with donor template) and imprecise genome editing, both of which have found applications in gene therapy, agriculture, and elsewhere. In addition, the nuclease-dead versions of Cas9 orthologs are being used for transcription modulation, site-specific DNA labeling, and for proteome profiling at specific genomic loci. Several different Cas9s have been used for these applications. Central to the programmability of Cas9 and hence its applications is the ability to

introduce any guide sequence in the crRNA. The crRNA and tracrRNA can be fused together to form a single-guide RNA (sgRNA), which is more stable and provides enhanced genome editing.

What is needed in the art are improved Cas9s and sgRNA sequences that can provide specific and accurate editing of a wider range of target sites, especially when combined with reliable nucleic acid delivery platforms.

Summary of The Invention

The present invention is related to compositions and methods for gene therapy. Several approaches described herein utilize *Neisseria meningitidis* Cas9 systems that provide hyperaccurate CRISPR gene editing platforms. Furthermore, the invention incorporates improvements of this Cas9 system: for example, truncating the single guide RNA sequences, and the packing of Nme1Cas9 or Nme2Cas9 with its guide RNA in an adeno-associated viral vector that is compatible for *in vivo* administration. Furthermore, Type II-C Cas9 orthologs have been identified that target protospacer adjacent motif sequences limited to between one – four required nucleotides.

In one embodiment, the present invention contemplates a single guide ribonucleic acid (sgRNA) sequence comprising a truncated repeat:anti-repeat region. In one embodiment, the sgRNA sequence further comprises a truncated Stem 2 region. In one embodiment, the sgRNA sequence further comprises a truncated spacer region. In one embodiment, said sgRNA sequence has a length of 121 nucleotides. In one embodiment, said sgRNA sequence length is selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides. In one embodiment, said sgRNA sequence has a length of 100 nucleotides. In one embodiment, said sgRNA sequence is an Nme1Cas9 single guide ribonucleic acid sequence. In one embodiment, said sgRNA sequence is an Nme2Cas9 single guide ribonucleic acid sequence. In one embodiment, said sgRNA sequence is an Nme1Cas9 single guide ribonucleic acid sequence or an Nme2Cas9 single guide ribonucleic acid sequence.

In one embodiment, the present invention contemplates a single guide ribonucleic acid (sgRNA) sequence comprising a truncated Stem 2 region. In one embodiment, the sgRNA sequence further comprises a truncated repeat:anti-repeat region. In one embodiment, the sgRNA sequence further comprises a truncated spacer region. In one embodiment, said sgRNA

sequence has a length is selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides. In one embodiment, said sgRNA sequence has a length of 100 nucleotides.

In one embodiment, the present invention contemplates an adeno-associated viral (AAV) vector comprising a single guide ribonucleic acid-*Neisseria meningitidis* Cas9 (sgRNA-Nme1Cas9 or sgRNA-Nme2Cas9) nucleic acid vector. In one embodiment, said single guide ribonucleic acid-*Neisseria meningitidis* Cas9 nucleic acid vector comprises at least one promoter. In one embodiment, said at least one promoter is selected from the group consisting of a U6 promoter and a U1a promoter. In one embodiment, said single guide ribonucleic acid-*Neisseria meningitidis* Cas9 nucleic acid vector comprises a Kozak sequence. In one embodiment, said sgRNA comprises a nucleic acid sequence that is complementary to a gene-of-interest sequence. In one embodiment, said gene-of-interest sequence is selected from the group consisting of a *PCSK9* sequence and a *ROSA26* sequence. In one embodiment, said sgRNA comprises an untruncated sequence that has a length of 145 nucleotides. In one embodiment, said sgRNA comprises a truncated repeat-antirepeat sequence. In one embodiment, said sgRNA further comprises a truncated Stem 2 region. In one embodiment, said sgRNA further comprises a truncated spacer region. In one embodiment, said sgRNA sequence has a length of 121 nucleotides. In one embodiment, said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides. In one embodiment, said sgRNA sequence has a length of 100 nucleotides. In one embodiment, said sgRNA comprises a truncated Stem 2 region. In one embodiment, said sgRNA further comprises a truncated repeat:antirepeat region. In one embodiment, said sgRNA further comprises a truncated spacer region. In one embodiment, said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 101 nucleotides, and 99 nucleotides. In one embodiment, said sgRNA sequence has a length of 100 nucleotides. In one embodiment, said sgRNA comprises an untruncated sequence has a length of 145 nucleotides.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a patient exhibiting at least one symptom of a medical condition, wherein said patient comprises a plurality of genes related to said medical condition; ii) a delivery platform comprising a single guide ribonucleic acid-*Neisseria meningitidis* Cas9 (sgRNA-Nme1Cas9 or

sgRNA-Nme2Cas9) nucleic acid vector, wherein said sgRNA comprises a nucleic acid sequence that is complementary to a portion of at least one of said plurality of genes; and b) administering said AAV plasmid to said patient under conditions such that said at least one symptom of said medical condition is reduced. In one embodiment, the delivery platform comprises an adeno-associated viral (AAV) vector. In one embodiment, the delivery platform comprises a microparticle. In one embodiment, said medical condition comprises hypercholesterolemia. In one embodiment, said medical condition comprises tyrosinemia. In one embodiment, said at least one of said plurality of genes is a *PCSK9* gene. In one embodiment, said sgRNA nucleic acid is complementary to a portion of said *PCSK9* gene. In one embodiment, at least one of said plurality of genes is an *FAH* gene. In one embodiment, said sgRNA nucleic acid is complementary to a portion of said *FAH* gene. In one embodiment, said sgRNA comprises a truncated repeat-antirepeat sequence. In one embodiment, said sgRNA further comprises a truncated Stem 2 region. In one embodiment, said sgRNA further comprises a truncated spacer region. In one embodiment, said sgRNA sequence has a length of 121 nucleotides. In one embodiment, said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 101 nucleotides, and 99 nucleotides. In one embodiment, said sgRNA sequence has a length of 100 nucleotides. In one embodiment, said sgRNA comprises a truncated Stem 2 region. In one embodiment, said sgRNA further comprises a truncated repeat:antirepeat region. In one embodiment, said sgRNA further comprises a truncated spacer region. In one embodiment, said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides. In one embodiment, said sgRNA sequence has a length of 100 nucleotides. In one embodiment, said sgRNA comprises an untruncated sequence has a length of 145 nucleotides.

In one embodiment, the present invention contemplates an adeno-associated viral (AAV) plasmid encoding a Type II-C Cas9 nuclease protein wherein said protein comprises a protospacer adjacent motif recognition domain configured with a binding site to a protospacer adjacent motif sequence comprising between one to four required nucleotides. In one embodiment, said Type II-C Cas9 nuclease protein is selected from the group consisting of a *Neisseria meningitidis strain De10444* Nme2Cas9 nuclease protein, a *Haemophilus parainfluenzae* HpaCas9 nuclease protein and a *Simonsiella muelleri* SmuCas9 nuclease protein.

In one embodiment, said protospacer adjacent motif sequence comprising one to four required nucleotides is selected from the group consisting of N_4CN_3, N_4CT, N_4CCN, N_4CCA, and N_4GNT_3. In one embodiment, the one to four required nucleotides are selected from the group consisting of C, CT, CCN, CCA, CN_3 and GNT_2. In one embodiment, said Type II-C Cas9 nuclease protein is bound to a truncated sgRNA. In one embodiment, the adeno-associated viral plasmid encodes two sgRNA sequences. In one embodiment, the adeno-associated viral plasmid encodes a poly-adenosine sequence. In one embodiment, the adeno-associated viral plasmid encodes a homology-directed repair donor nucleotide template. In one embodiment, the adeno-associated viral plasmid is an all-in-one adeno-associated viral plasmid.

In one embodiment, the present invention contemplates, a method, comprising: a) providing; i) a patient exhibiting at least one symptom of a medical condition, wherein said patient comprises a plurality of genes related to said medical condition, wherein said plurality of genes comprise a protospacer adjacent motif comprising between one – four required nucleotides; ii) a delivery platform comprising at least one nucleic acid encoding a Type II-C Cas9 nuclease protein wherein said protein comprises a protospacer adjacent motif recognition domain configured with a binding site to said protospacer adjacent motif sequence comprising between two – four required nucleotides; and b) administering said delivery platform to said patient under conditions such that said at least one symptom of said medical condition is reduced. In one embodiment, said medical condition comprises hypercholesterolemia. In one embodiment, said medical condition comprises tyrosinemia. In one embodiment, said at least one of said plurality of genes is a *PCSK9* gene. In one embodiment, said sgRNA nucleic acid is complementary to a portion of said *PCSK9* gene. In one embodiment, at least one of said plurality of genes is an *FAH* gene. In one embodiment, said sgRNA nucleic acid is complementary to a portion of said *FAH* gene. In one embodiment, said delivery platform comprises an adeno-associated viral plasmid. In one embodiment, said delivery platform comprises a microparticle. In one embodiment, said Type II-C Cas9 nuclease protein is selected from the group consisting of a *Neisseria meningitidis strain De10444* Nme2Cas9 nuclease protein, a *Haemophilus parainfluenzae* HpaCas9 nuclease protein and a *Simonsiella muelleri* SmuCas9 nuclease protein. In one embodiment, said protospacer adjacent motif sequence comprising one - four required nucleotides is selected from the group consisting of N_4CN_3, N_4CT, N_4CCN, N_4CCA, and N_4GNT_3. In one embodiment, the one to four required nucleotides

are selected from the group consisting of C, CT, CCN, CCA, CN_3 and GNT_2. In one embodiment, said Type II-C Cas9 nuclease protein is bound to a truncated sgRNA. In one embodiment, the adeno-associated viral plasmid encodes two sgRNA sequences. In one embodiment, the adeno-associated viral plasmid encodes a poly-adenosine sequence. In one embodiment, the adeno-associated viral plasmid encodes a homology-directed repair donor nucleotide template. In one embodiment, the adeno-associated viral plasmid is an all-in-one adeno-associated viral plasmid.

In one embodiment, the present invention contemplates an adeno-associated viral (AAV) plasmid encoding a Type II-C Cas9 nuclease protein wherein said protein comprises a protospacer adjacent motif recognition domain (e.g., a PAM-Interacting Domain; PID) configured to bind with a protospacer adjacent motif (PAM) sequence, said PAM sequence comprising an adjacent cytosine dinucleotide pair. In one embodiment the adjacent cytosine dinucleotide pair is at the PAM positions five (5) and six (6). In one embodiment, said Type II-C Cas9 nuclease protein is derived from a *Neisseria meningitidis* strain. In one embodiment, the *Neisseria meningitidis* strain is De10444. In one embodiment, the Type II-C Cas9 nuclease protein is an Nme2Cas9 nuclease protein. In one embodiment, the *Neisseria meningitidis* strain is 98002. In one embodiment, the Type II-C Cas9 nuclease protein is an Nme3Cas9 nuclease protein. In one embodiment, said PAM sequence is selected from the group consisting of N_4CC, N_4CCN_3, N_4CCA, $N_4CC(X)$, N_4CA_3 and N_{10}. In one embodiment, the PAM sequence is N_3CC. In one embodiment, the Type II-C Cas9 nuclease protein further comprises an sgRNA sequence. In one embodiment, the sgRNA sequence comprises a spacer ranging in length between approximately seventeen (17) – twenty four (24) nucleotides.

In one embodiment, the present invention contemplates a method, comprising: a) providing; i) a patient exhibiting at least one symptom of a medical condition, wherein said patient comprises a plurality of genes related to said medical condition, wherein said plurality of genes comprise a protospacer adjacent motif comprising an adjacent cytosine dinucleotide pair; ii) a delivery platform comprising at least one nucleic acid encoding a Type II-C Cas9 nuclease protein wherein said protein comprises a protospacer adjacent motif recognition domain (e.g., a PAM Interacting Domain; PID) configured to bind with said protospacer adjacent motif sequence comprising an adjacent cytosine dinucleotide pair; and b) administering said delivery platform to said patient under conditions such that said at least one symptom of said medical

condition is reduced. In one embodiment, said delivery platform comprises an adeno-associated viral vector. In one embodiment, the adeno-associated viral vector is adeno-associated viral vector eight (AAV8). In one embodiment, said medical condition comprises hypercholesterolemia. In one embodiment, said medical condition comprises tyrosinemia. In one embodiment, the medical condition is x-linked chronic granulomatous disease. In one embodiment, the medical condition is aspartylglycosaminuria. In one embodiment, said at least one of said plurality of genes is a *PCSK9* gene. In one embodiment, said sgRNA nucleic acid is complementary to a portion of said *PCSK9* gene. In one embodiment, at least one of said plurality of genes is an *FAH* gene. In one embodiment, said sgRNA nucleic acid is complementary to a portion of said *FAH* gene. In one embodiment, the adeno-associated viral plasmid encodes at least one sgRNA sequence. In one embodiment, the adeno-associated viral plasmid encodes two sgRNA sequences. In one embodiment, the adeno-associated viral plasmid encodes a poly-adenosine sequence. In one embodiment, the adeno-associated viral plasmid encodes a homology-directed repair donor nucleotide template. In one embodiment, the adeno-associated viral plasmid is an all-in-one adeno-associated viral plasmid. In one embodiment, said delivery platform comprises a microparticle. In one embodiment the adjacent cytosine dinucleotide pair is at the PAM positions five (5) and six (6). In one embodiment, said Type II-C Cas9 nuclease protein is derived from a *Neisseria meningitidis* strain. In one embodiment, the *Neisseria meningitidis* strain is *De10444*. In one embodiment, the Type II-C Cas9 nuclease protein is an Nme2Cas9 nuclease protein. In one embodiment, the *Neisseria meningitidis* strain is 98002. In one embodiment, the Type II-C Cas9 nuclease protein is an Nme3Cas9 nuclease protein. In one embodiment, said PAM sequence is selected from the group consisting of N_4CC, N_4CCN_3, N_4CCA, $N_4CC(X)$, N_4CA_3 and N_{10}. In one embodiment, the PAM sequence is N_3CC. In one embodiment, the Type II-C Cas9 nuclease protein further comprises an sgRNA sequence. In one embodiment, the sgRNA sequence comprises a spacer ranging in length between approximately seventeen (17) – twenty four (24) nucleotides.

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 also plural entities and also includes 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.

The term "about" or "approximately" as used herein, in the context of any of any assay measurements refers to +/- 5% of a given measurement.

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" sequence. 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. Doudna et al. Genome editing. The new frontier of genome engineering with CRISPR-Cas9" *Science* 346(6213):1258096 (2014).

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. tracrRNA and spacer RNA may be combined 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, Jinek et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity" *Science* 337(6096):816-821 (2012).

As used herein, the term "catalytically active Cas9" refers to an unmodified Cas9 nuclease comprising full nuclease activity.

The term "nickase" as used herein, refers to a nuclease that cleaves only a single DNA strand, either due to its natural function or because it has been engineered to cleave only a single DNA strand. Cas9 nickase variants that have either the RuvC or the HNH domain mutated

provide control over which DNA strand is cleaved and which remains intact. Jinek et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity" *Science* 337(6096):816-821 (2012) and Cong et al. Multiplex genome engineering using CRISPR/Cas systems" Science 339(6121):819-823 (2013).

The term, "trans-activating crRNA", "tracrRNA" as used herein, refers to a small trans-encoded RNA. For example, CRISPR/Cas (clustered, regularly interspaced short palindromic repeats/CRISPR-associated proteins) constitutes an RNA-mediated defense system, which protects against viruses and plasmids. This defensive pathway has three steps. First a copy of the invading nucleic acid is integrated into the CRISPR locus. Next, CRISPR RNAs (crRNAs) are transcribed from this CRISPR locus. The crRNAs are then incorporated into effector complexes, where the crRNA guides the complex to the invading nucleic acid and the Cas proteins degrade this nucleic acid. There are several pathways of CRISPR activation, one of which requires a tracrRNA, which plays a role in the maturation of crRNA. TracrRNA is complementary to the repeat sequence of the pre-crRNA, forming an RNA duplex. This is cleaved by RNase III, an RNA-specific ribonuclease, to form a crRNA/tracrRNA hybrid. This hybrid acts as a guide for the endonuclease Cas9, which cleaves the invading nucleic acid.

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).

The terms "protospacer adjacent motif recognition domain", "PAM Interacting Domain" or "PID" as used herein, refers to a Cas9 amino acid sequence that comprises a binding site to a DNA target PAM sequence.

The term "binding site" as used herein, refers to any molecular arrangement having a specific tertiary and/or quaternary structure that undergoes a physical attachment or close association with a binding component. For example, the molecular arrangement may comprise a sequence of amino acids. Alternatively, the molecular arrangement may comprise a sequence a nucleic acids. Furthermore, the molecular arrangment may comprise a lipid bilayer or other biological material.

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 "orthogonal" refers to targets that are non-overlapping, uncorrelated, or independent. For example, if two orthogonal Cas9 isoforms were utilized, they would employ orthogonal sgRNAs that only program one of the Cas9 isoforms for DNA recognition and cleavage. Esvelt et al., "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing" *Nat Methods* 10(11):1116-1121 (2013). For example, this would allow one Cas9 isoform (e.g. *S. pyogenes* Cas9 or SpyCas9) to function as a nuclease programmed by a sgRNA that may be specific to it, and another Cas9 isoform (e.g. *N. meningitidis* Cas9 or NmeCas9) to operate as a nuclease-dead Cas9 that provides DNA targeting to a binding site through its PAM specificity and orthogonal sgRNA. Other Cas9s include *S. aureus* Cas9 or SauCas9 and *A. naeslundii* Cas9 or AnaCas9.

The term "truncated" as used herein, when used in reference to either a polynucleotide sequence or an amino acid sequence means that at least a portion of the wild type sequence may be absent. In some cases, truncated guide sequences within the sgRNA or crRNA may improve the editing precision of Cas9. Fu, et al. "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs" Nat Biotechnol. 2014 Mar;32(3):279-284 (2014).

The term "base pairs" as used herein, refer to specific nucleobases (also termed nitrogenous bases), that are the building blocks of nucleotide sequences that form a primary structure of both DNA and RNA. Double-stranded DNA may be characterized by specific hydrogen bonding patterns. Base pairs may include, but are not limited to, guanine-cytosine and adenine-thymine base pairs.

The term "specific genomic target" as used herein, refers to any pre-determined nucleotide sequence capable of binding to a Cas9 protein contemplated herein. The target may include, but may be not limited to, a nucleotide sequence complementary to a programmable DNA binding domain or an orthogonal Cas9 protein programmed with its own guide RNA, a

nucleotide sequence complementary to a single guide RNA, a protospacer adjacent motif recognition sequence, an on-target binding sequence and an off-target binding sequence.

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 "fails to bind" as used herein, refers to any nucleotide-nucleotide interaction or a nucleotide-amino acid interaction that exhibits partial complementarity, but has insufficient complementarity for recognition to trigger the cleavage of the target site by the Cas9 nuclease. Such binding failure may result in weak or partial binding of two molecules such that an expected biological function (e.g., nuclease activity) fails.

The term "cleavage" as used herein, may be defined as the generation of a break in the DNA. This could be either a single-stranded break or a double-stranded break depending on the type of nuclease that may be employed.

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 or the specific inclusion of new sequence through the use of an exogenously supplied DNA template. Such a specific genomic target includes, but may be not limited to, a chromosomal region, mitochondrial DNA, a gene, a promoter, an open reading frame or any nucleic acid sequence.

The term "delete", "deleted", "deleting" or "deletion" as used herein, may be defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are, or become, absent.

The term "gene of interest" as used herein, refers to any pre-determined gene for which deletion may be desired.

The term "allele" as used herein, refers to any one of a number of alternative forms of the same gene or same genetic locus.

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 “derived from” as used herein, refers to the source of a compound or sequence. In one respect, a compound or sequence may be derived from an organism or particular species. In another respect, a compound or sequence may be derived from a larger complex or sequence.

The term “protein” as used herein, refers to any of numerous naturally occurring extremely complex substances (as an enzyme or antibody) that consist of amino acid residues joined by peptide bonds, contain the elements carbon, hydrogen, nitrogen, oxygen, usually sulfur. In general, a protein comprises amino acids having an order of magnitude within the hundreds.

The term “peptide” as used herein, refers to any of various amides that are derived from two or more amino acids by combination of the amino group of one acid with the carboxyl group of another and are usually obtained by partial hydrolysis of proteins. In general, a peptide comprises amino acids having an order of magnitude with the tens.

The term "polypeptide", refers to any of various amides that are derived from two or more amino acids by combination of the amino group of one acid with the carboxyl group of another and are usually obtained by partial hydrolysis of proteins. In general, a peptide comprises amino acids having an order of magnitude with the tens or larger.

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.

"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.

The term "biologically active" refers to any molecule having structural, regulatory or biochemical functions. For example, biological activity may be determined, for example, by restoration of wild-type growth in cells lacking protein activity. Cells lacking protein activity may be produced by many methods (i.e., for example, point mutation and frame-shift mutation). Complementation is achieved by transfecting cells which lack protein activity with an expression vector which expresses the protein, a derivative thereof, or a portion thereof.

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.

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)).

Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription. Maniatis, T. et al., *Science* 236:1237 (1987). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in plant, yeast, insect and mammalian cells and viruses (analogous control elements, i.e., promoters, are also found in prokaryotes). The selection of a particular promoter and enhancer depends on what cell type is to be used to express the protein of interest.

The term "poly A site" or "poly A sequence" as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable as transcripts lacking a poly A tail are unstable and are rapidly degraded. The poly A signal utilized in an expression vector may be "heterologous" or "endogenous." An endogenous poly A signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly A signal is one which is isolated from one gene and placed 3' of another gene. Efficient expression of recombinant DNA sequences in eukaryotic cells involves expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length.

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 "coding region" when used in reference to a structural gene refers to the nucleotide sequences which encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eukaryotes, on the 5' side by the nucleotide triplet "ATG" which encodes the initiator methionine and on the 3' side by one of the three triplets which specify stop codons (i.e., TAA, TAG, TGA).

As used herein, the term "structural gene" refers to a DNA sequence coding for RNA or a protein. In contrast, "regulatory genes" are structural genes which encode products which control the expression of other genes (e.g., transcription factors).

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 "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.

Brief Description Of The Figures

The file of this patent contains at least one drawing executed in color. Copies of this patent with color drawings will be provided by the Patent and Trademark Office upon request and payment of the necessary fee.

Figure 1 presents representative sequence of a conventional, full-length, 145 nt Nme1Cas9 and Nme2Cas9 sgRNA.

Figure 2 presents exemplary Nme1Cas9 sgRNA sequences and associated gene editing activity having a truncated repeat:anti-repeat region or a truncated Stem 2 region. Deletion/truncation series of Nme1Cas9 sgRNAs. Top: aligned sequences, color-coded as in

Figure 1. Bottom: T7E1 assays of editing at Nme1Cas9 target site 7 (NTS7), using the indicated sgRNAs as guides.

Figure 3 presents exemplary Nme1Cas9 sgRNA sequences and associated gene editing activity having a truncated repeat:anti-repeat region or a truncated Stem 2 region. The shortest Nme1Cas9 sgRNAs (#10 - 101 nt; 24 nt guide sequence; and #11 - 100 nt; 23 nt guide sequence) efficiently edit three distinct target sites (NTS7, NTS27, and NTS55) in the human genome. Top: sequences of wild-type and minimized sgRNAs, using the same color scheme as in the previous figures. Bottom: T7E1 assays of editing efficiency at the three target sites in HEK293T cells.

Figure 4 presents exemplary sequences (as secondary structures) of Nme1Cas9 wt sgRNA, and truncated sgRNAs 11 and 12 and associated gene editing by RNP delivery of Nme1Cas9 and sgRNAs. Three genomic sites (N-TS72, N-TS55 and N-TS40), and one traffic light reporter site was targeted in the human genome using HEK293T cells. Top: sequences shown as secondary structures of wild-type and minimized sgRNAs. Bottom: Editing efficiencies measured by T7E1 assay or flow cytometry are depicted as bar graphs.

Figure 5 presents gene editing in PLB985 cells using minimized sgRNA 11, and in vitro transcribed wt sgRNA. Cells were transfected with RNP complexes of Nme1Cas9 and sgRNAs and gene editing at genomic site N-TS72 measured by TIDE.

Figure 6 presents a schematic of one embodiment of an AAV vector comprising a complete CRISPR/Cas9 gene editing complex. Representative sequences of the various AAV vector regions are color coded in Appendix 1.

Figure 7 presents one embodiment of a color-coded sequence of Nme single-guide RNA and a promoter as depicted in Figure 4, wherein the backbone is linearized using SapI to insert a 24-nt target spacer.

U6 promoter: Turquoise.

Nme single guide RNA: Purple

SapI restriction sites: Bold

Figure 8 presents one embodiment of a color-coded sequence of an Nme1Cas9 and promoter as depicted in Figure 4, wherein Start and Stop codons underlined in bold.

U1a promoter: Blue

Kozak sequence: Grey

Humanized Nme1Cas9: Red

SV40 NLS: Green

Nucleoplasmin (NP) NLS: Yellow

HA Tags (3X): Bold Orange

Synthetic NLS: Turquoise

Beta-globin polyadenylation signal: Teal

Figure 9 presents exemplary data showing editing efficiency of various target sites using AAV plasmids with sgRNA-Nme1Cas9 constructs guided to either the *Pcsk9* gene or the *Rosa26* gene (control).

Figure 10 presents one embodiment of color-coded target site sequences for sgRNA-Nme1Cas9 constructs guided to either a *Pcsk9* gene or a *Rosa26* gene (control).

24-nt Nme1Cas9 target spacer, blue bold

Nme1Cas9 PAM underlined [NNNNGATT)

T7E1 primers binding sites: green italics

TIDE primers binding sites: purple italics

Figure 11 presents exemplary data showing gene editing efficiency following in vivo hydrodynamic injection by mouse tail vein of 30 μg of endotoxin-free sgRNA-Nme1Cas9-AAV plasmid targeting *Pcsk9*.

Figure 12A presents exemplary data showing gene editing efficiency in the liver at the *Pcsk9* gene and the *Rosa26* gene by Nme1-Cas9 vector packaged in hepatocyte-specific AAV8 serotype, at a dose of $4x10^{11}$ genomic copies (gc) per mouse 14 days post vector administration. Figure 12B presents exemplary data showing gene editing efficiency in the liver at a *Pcsk9* gene and a *Rosa26* gene by an Nme1-Cas9 vector packaged in hepatocyte-specific AAV8 serotype, at a dose of $4x10^{11}$ genomic copies (gc) per mouse 50 days post vector administration.

Figure 13 presents exemplary data showing reduction in mouse cholesterol levels following injection of sgRNA-Cas9-AAV vectors targeting a *Pcsk9* gene, a *Rosa26* gene and a PBS control group at 0, 25 and 50 days.

Figures 14A and 14B present exemplary data showing a genome-wide unbiased identification of double strand breaks (DSBs) enabled by sequencing (e.g., GUIDE-Seq) assay that searched for off-target editing sites for both the *Pcsk9*-sgRNA-Cas9-AAV (A) and the *Rosa26*-sgRNA-Cas9-AAV (B).

Figure 15 presents exemplary data showing a targeted TIDE analyses in mice 14 days post-injection of both the *Pcsk9*-sgRNA-Cas9-AAV and the *Rosa26*-sgRNA-Cas9-AAV that revealed minimal cleavage. OnT, on-target site; OT1, OT2 etc.: off-target sites.

Figure 16 presents exemplary data showing a hematoxylin and eosin stain assay in the liver sections of mice sacrificed at day 14 subsequent to injection of vectors targeting a *Pcsk9* gene and a *Rosa26* gene. No evidence for a host immune response is observed.

Figure 17 illustrates one embodiment of an in vitro PAM library identification workflow. NGS, next-generation sequencing.

Figure 18 presents putative sequence from an in vitro PAM discovery assay depicted in Figure 17. Recombinantly purified Cas9 from each bacterium was incubated with an sgRNA and a target with randomized PAM. Nme1Cas9 was used as a control.

Figure 19 presents exemplary data showing percent genome editing at a single site (top panel) in the human genome in HEK293T cells. Percentages show estimated indel formation using a T7E1 endonuclease assay (Nme2Cas9, HpaCas9) or a fluorescent assay (for SmuCas9) based on the "traffic light" reporter integrated into the genome of HEK293T cells.

Figure 20 presents exemplary data showing genome editing in HEK293T cells of an integrated traffic light reporter with Nme2Cas9 targeting various protospacers with various PAMs (X-axis). The results suggest a preferred NNNNCC PAM for Nme2Cas9 in human cells.

Figure 21 presents exemplary data showing genome editing in HEK293T cells in the presence of various anti-CRISPR (Acr) proteins. T7E1 digestion shows genome editing following plasmid transfection (to express Nme2Cas9 and its sgRNA) or RNA/protein delivery (HpaCas9 and its sgRNA). Nme2Cas9 is robustly inhibited by two Acr proteins ($AcrIIC3_{Nme}$ and $AcrIIC4_{Hpa}$), while HpaCas9 is inhibited by four of the previously reported type II-C Acrs. These results show that these two Cas9 proteins are subject to off-switch control by anti-CRISPRs.

Figure 22 presents exemplary data of traffic light reporter (TLR) gene editing using the Nme2Cas9-sgRNA complex on "CC" dinucleotide PAMs. Figure 22A. Blue bars are the % of cells that exhibit fluorescence, whereas red bars indicate % editing more accurately based on sequencing ("TIDE analysis").

Figure 23 presents exemplary data of gene editing by Nme2Cas9 using T7E1 assays at the *AAVS1*, Chromosome 14 NTS4, *VEGF* and *CFTR* loci.

Figure 24 presents one embodiment for a wild type Nme2Cas9 bacterial open reading frame DNA sequence.

Figure 25 presents one embodiment of a wild type Nme2Cas9 bacterial protein sequence.

Figure 26 presents one embodiment of an Nme2Cas9 human-codon-optimized open reading frame DNA sequence. Yellow - SV40 NLS; Green - 3X-HA-Tag; Blue: cMyc-like NLS.

Figure 27 presents one embodiment of an Nme2Cas9 humanized protein sequence. Yellow - SV40 NLS; Green - 3X-HA-Tag; Blue: cMyc-like NLS.

Figure 28 presents one embodiment of an HpaCas9 bacterial protein sequence.

Figure 29 presents one embodiment of an SmuCas9 native bacterial open reading frame DNA sequence.

Figure 30 presents one embodiment of an SmuCas9 bacterial protein sequence.

Figure 31 presents one embodiment of an SmuCas9 Human-codon-optimized open reading frame DNA sequence. Yellow - SV40 NLS; Green - 3X-HA-Tag; Blue: cMyc-like NLS.

Figure 32 presents one embodiment of an SmuCas9 humanized protein sequence. Yellow - SV40 NLS; Green - 3X-HA-Tag; Blue: cMyc-like NLS.

Figure 33 presents exemplary Type-II C Cas9 ortholog single guide RNA sequences compatible with short C-rich PAMs. Yelllow - crRNA; Gray – Linker; Purple – tracrRNA.

Figure 34 illustrates three closely related *Neisseria meningitidis* Cas9 orthologs that have distinct PAMs.

Figure 34A: 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 34B: Experimental workflow of the *in vitro* PAM discovery assay with a 10 nt randomized PAM sequence downstream of a protospacer. Adapters were ligated to cleaved product and sequenced.

Figure 34C: Sequence logos of the *in vitro* PAM discovery assay demonstrating an N_4GATT PAM for Nme1Cas9, as shown previously in cells.

Figure 34D: Sequence logos showing Nme1Cas9 with its PID swapped with those of Nme2Cas9 (left) and Nme3Cas9 (right) recognize a C at position

5. The remaining nucleotides were determined with lower confidence due to the modest cleavage efficiency of the protein chimeras (Figure 35C).

Figure 34E: Sequence logo illustrating that full-length Nme2Cas9 recognizes an N_4CC PAM based on the PAM discovery assay with a fixed C at position 5, and PAM nts 1-4 and 6-8 randomized.

Figure 35 shows a characterization of *Neisseria meningitidis* Cas9 orthologs with rapidly-evolving PIDs in accordance with Figure 34.

Figure 35A: 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. Group 1 (blue) PIDs with >98% identity to Nme1Cas9, group 2 (orange) with PIDs ~52% identical to Nme1Cas9, and group 3 (green) with PIDs ~86% identical to Nme1Cas9. Three representative Cas9 orthologs from each group (Nme1Cas9, Nme2Cas9 and Nme3Cas9) are marked.

Figure 35B: Schematic showing the CRISPR loci of the strains encoding the three Cas9 orthologs (Nme1Cas9, Nme2Cas9, and Nme3Cas9) from (A). Percent identities of each CRISPR-Cas component to *N. meningitidis* 8013 (encoding Nme1Cas9) are shown.

Figure 35C: Number of reads from cleaved DNAs from the *in vitro* assays for intact Nme1Cas9, and for chimeras with Nme1Cas9's PID swapped with those of Nme2Cas9 and Nme3Cas9. The reduced read counts indicate lower cleavage efficiencies in the chimeras.

Figure 35D: Sequence logos from the *in vitro* PAM discovery assay on an NNNNCNNN randomized PAM by Nme1Cas9 with its PID swapped with those of Nme2Cas9 (left) or Nme3Cas9 (right).

Figure 36 shows that the Nme2Cas9 uses a 22-24 nucleotide spacer to recognize and edit sites adjacent to an N_4CC PAM. All experiments were done in triplicate, and error bars represent standard error of mean (s.e.m.).

Figure 36A: Schematic showing the transient transfection workflow on HEK293T TLR2.0 cells. Nme2Cas9 and sgRNA plasmids were transfected and mCherry+ cells were detected 72 hours after transfection.

Figure 36B: Using Nme2Cas9 to target an array of PAMs in TLR2.0. All sites with N_4CC PAMs were targeted with varying degrees of efficiency, while no Nme2Cas9 targeting observed at an N_4GATT PAM or in the absence of sgRNA. SpyCas9 (targeting NGG) and Nme1Cas9 (targeting N_4GATT) were used as positive controls.

Figure 36C: The effect of spacer length on the efficiency of Nme2Cas9 editing. An sgRNA targeting a TLR2.0 site (with an N_4CCA PAM) with spacer lengths varying from 24 to 20 nts (including a 5'-terminal G), showing highest editing efficiencies with 22-24 nucleotide spacers.

Figure 36D: Nme2Cas9 nickases (HNH nickase = $Nme2Cas9^{D16A}$; RuvC nickase = $Nme2Cas9^{H588A}$) can be used in tandem to generate indels in TLR2.0. Targets with cleavage sites 32 base pairs and 64 base pairs apart were targeted using either nickase to generate indels. The HNH nickase shows efficient editing, particularly when the cleavage sites were close (32 bp). Wildtype Nme2Cas9 was used as a control. Green is GFP (HDR) and red is mCherry (NHEJ).

Figure 37 presents exemplary data regarding PAM, spacer, and seed elements for Nme2Cas9 targeting in mammalian cells, in accordance with Figure 36. All experiments were done in triplicate and error bars represent s.e.m.

Figure 37A: Nme2Cas9 targeting at N_4CD sites in TLR2.0. Four sites for each non-C nucleotide at the tested position (N_4CA, N_4CT and N_4CG) were examined, and an N_4CC site was used as a positive control.

Figure 37B: Nme2Cas9 targeting at N_4DC sites in TLR2.0 [similar to (A)].

Figure 37C: Guide truncations on another TLR2.0 site, revealing similar length requirements as those observed in Figure 36C.

Figure 37D: Nme2Cas9 targeting efficiency is differentially sensitive to single-nucleotide mismatches in the seed sequence. Data show the effects of walking single-nucleotide mismatches in the sgRNA along the 23-nt spacer in a TLR target site.

Figure 38 presents exemplary data showing Nme2Cas9 genome editing efficiency at genomic loci in mammalian cells via multiple delivery methods. All results represent 3 independent biological replicates, and error bars represent s.e.m.

Figure 38A: Nme2Cas9 genome editing using transient transfections with sgRNAs targeting loci throughout the human genome in HEK293T cells. 14 sites were selected based the initial screening of 38 sites to demonstrate the range of indels (as detected by TIDE) at different loci induced by Nme2Cas9. An Nme1Cas9 target site (with an N_4GATT PAM) was used as a negative control.

Figure 38B: Left panel: Transient transfection of an all-in-one plasmid (Nme2Cas9 + sgRNA) targeting the *Pcsk9* and *Rosa26* loci 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 at the intended loci.

Figure 38C: Nme2Cas9 can be electroporated as an RNP complex for efficient 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 using TIDE after 72h.

Figure 39 presents exemplary data showing dose dependence and block deletions by Nme2Cas9, in accordance with Figure 38.

Figure 39A: Increasing the dose of electroporated Nme2Cas9 plasmid (500 ng, vs. 200 ng in Fig. 3A) improves editing efficiency at two sites (TS16 and TS6).

Figure 39B: Nme2Cas9 can be used to create block deletions. Two TLR2.0 targets with cleavage sites 32 bp apart were targeted simultaneously with Nme2Cas9. The majority of lesions created were exactly 32 bp deletions (green).

Figure 40 presents exemplary data showing that Type II-C Anti-CRISPR proteins can be used to inhibit Nme2Cas9 gene editing acitivity (e.g., as an off-switch) *in vitro* and *in vivo*. All experiments were done in triplicate and error bars represent s.e.m.

Figure 40A: *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 control Acr (AcrE2). All other previously characterized Acrs inhibited Nme1Cas9, as expected. Bottom: Nme2Cas9 efficiently cleaves a target containing a protospacer with an N_4CC PAM in the presence of AcrE2 and and AcrIIC5$_{Smu}$, suggesting that AcrIIC5$_{Smu}$ is unable to inhibit Nme2Cas9 at a 10:1 molar ratio.

Figure 40B: Genome editing in the presence of the five previously described anti-CRISPR proteins. Plasmids expressing Nme2Cas9, sgRNA and each respective Acr (200 ng Cas9, 100 ng sgRNA, 200 ng Acr) were co-transfected into HEK293T cells, and genome editing was measured using TIDE 72 hr post transfection. Except for AcrE2 and AcrIIC5$_{Smu}$, all other Acrs inhibited genome editing, albeit at different efficiencies.

Figure 40C: Acr inhibition of Nme2Cas9 is dose-dependent with distinct apparent potencies. AcrIIC1$_{Nme}$ and AcrIIC4$_{Hpa}$ inhibit Nme2Cas9 completely at 2:1 and 1:1 ratios of cotransfected plasmids, respectively.

Figure 41 presents exemplary data showing that a Nme2Cas9 PID swap renders Nme1Cas9 insensitive to AcrIIC5$_{Smu}$ inhibition, in accordance with Figure 40. *In vitro* cleavage by the Nme1Cas9-Nme2Cas9PID chimera was performed in the presence of previously characterized Acr proteins (10 uM Cas9-sgRNA + 100 uM Acr).

Figure 42 presents exemplary data showing that Nme2Cas9 has no detectable off-targets in mammalian cells.

Figure 42A: Schematic showing the dual sites (DS) targetable by both SpyCas9 and Nme2Cas9 by virtue of their non-overlapping PAMs. The Nme2Cas9 PAM (orange) and SpyCas9 PAM (blue) are highlighted.

Figure 42B: Nme2Cas9 and SpyCas9 induce indels at dual sites. Six dual sites in *VEGFA* with GN$_3$GN$_{19}$NGGNCC sequences were selected for direct comparisons between the two orthologs. Plasmids expressing each Cas9 (with same promoter and NLSs) were transfected along with each

ortholog's cognate guide in HEK293T cells. Indel rates were determined by TIDE 72 hrs post transfection. Nme2Cas9 editing was detectable at all six sites and was more efficient than SpyCas9 on two sites (DS2 and 6). SpyCas9 edited four out of six sites (DS1, 2, 4 and 6), with two sites showing significantly higher editing rates than Nme2Cas9 (DS1 and 4). DS2, 4 and 6 were selected for GUIDE-Seq analysis as Nme2Cas9 was equally efficient, less efficient and more efficient than SpyCas9 at these sites, respectively.

Figure 42C: Nme2Cas9 has a clean off-target profile in human cells. Numbers of off-target sites detected by GUIDE-Seq for each nuclease at individual target sites are shown. SpyCas9 off-target numbers are shown in black. In addition to dual sites, TS6 (because of its high efficiency and potential for off-targets) and two mouse sites (to test accuracy in another cell type) also showed zero or one off-target site per guide.

Figure 42D: Targeted deep sequencing confirms the high Nme2Cas9 accuracy indicated by GUIDE-seq. Top off-target loci detected by GUIDE-seq were amplified and deep-sequenced. SpyCas9 showed off-targeting at most loci, while for Nme2Cas9, only one (the *Rosa26* site) showed editing at the off-target locus at relatively low levels (~40% on-target vs ~1% off-target). Note the log scale on the *y* axis.

Figure 42E: Nme2Cas9 and SpyCas9 efficiencies vary based on the locus and target site. Sites throughout the genome (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. Box-and-whisker plots indicate editing efficiencies at twenty-eight (28) dual sites by Nme2Cas9 and SpyCas9 (left). The sites that showed no editing were excluded from the analysis. Relative efficiencies of Nme2Cas9 and SpyCas9 show that Nme2Cas9 is less efficient than SpyCas9 (right), on average. Editing efficiencies by both

Cas9 orthologs at all twenty-eight (28) sites were included in the analysis of relative efficiencies in the right panel.

Figure 42F presents nucleic acids sequences 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 43 presents exemplary data showing the orthogonality and relative accuracy of Nme2Cas9 and SpyCas9 at dual target sites, in accordance with Figure 42.

Figure 43A: Nme2Cas9 and SpyCas9 guides are orthogonal. TIDE results show the frequencies of indels created by both nucleases targeting DS12 with either their cognate sgRNAs, or with the sgRNAs of the other ortholog.

Figure 43B: Nme2Cas9 and SpyCas9 exhibit comparable on-target editing efficiencies during 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 43C: SpyCas9's on-target vs. off-target reads for each site. Orange bars represent the on-target reads while black bars represent off-targets.

Figure 43D: Nme2Cas9's on-target vs off-target reads for each site.

Figure 43E: Bar graphs showing TIDE at expected off-target sites based on CRISPRseek, detecting no indels at off-target loci.

Figure 44 presents exemplary data showing Nme2Cas9 genome editing *in vivo* via all-in-one AAV delivery.

Figure 44A: Workflow for delivery of AAV8.Nme2Cas9+sgRNA to lower cholesterol levels in mice by targeting *Pcsk9*. Top: schematic of the all-in-one AAV vector expressing Nme2Cas9 and the sgRNA. Bottom: Timeline for AAV8.Nme2Cas9+sgRNA tail-vein injections, followed by cholesterol measurements at day 14 and indel, histology and cholesterol analyses at day 28.

Figure 44B: Deep sequencing analysis to measure indels in DNA extracted from livers of mice injected with AAV8.Nme2Cas9+sgRNA targeting *Pcsk9* and *Rosa26* (control) loci.

Figure 44C: Reduced serum cholesterol levels in mice injected with the *Pcsk9*-targeting guide compared to the *Rosa26*-targeting controls. *P* values are calculated by unpaired T-test.

Figure 44D: H&E staining from livers of mice injected with AAV8.Nme2Cas9+sgRosa26 (left) or AAV8.Nme2Cas9+sgPcsk9 (right) vectors. Scale bar, 25 um.

Figure 45 presents one embodiment of minimized AAV backbone and exemplary comparative TLR 2.0 data to the conventional sized AAV backbone.

Figure 46 presents a comparison of Nme2Cas9 structures of truncated sgRNA 11 with truncated sgRNA 12.

Figure 47 illustrates one embodiment of a minimized all-in-one AAV with a short polyA signal.

Figure 48 illustrates two embodiments of a minimized all-in-one AAV backbone. Dual sgRNAs in tandem (Top). Donor template for homology directed repair (Bottom).

Figure 49 presents a validation of an all-in-one AAV-sgRNA-hNme1Cas9 construct.

Figure 49A: Schematic representation of a single rAAV vector expressing human-codon optimized Nme1Cas9 and its sgRNA. The backbone is flanked by AAV inverted terminal repeats (ITR). The poly(a) signal is from rabbit beta-globin (BGH).

Figure 49B: Schematic diagram of the *Pcsk9* (top) and *Rosa26* (bottom) mouse genes. Red bars represent exons. Zoomed-in views show the protospacer sequence (red) whereas the Nme1Cas9 PAM sequence is highlighted in green. Double-stranded break location sites are denoted (black arrowheads).

Figure 49C: Stacked histogram showing a representative percentage distribution of insertions-deletions (indels) obtained by TIDE after AAV-sgRNA-hNme1Cas9 plasmid transfections in Hepa1–6 cells targeting *Pcsk9* (sgPcsk9) and *Rosa26* (sgRosa26) genes. Data are presented as mean values ± SD from three biological replicates.

Figure 49D: Stacked histogram showing a representative percentage distribution of indels at *Pcsk9* in the liver of C57Bl/6 mice obtained by TIDE after hydrodynamic injection of AAV-sgRNA-hNme1Cas9 plasmid.

Figure 50 presents exemplary data showing that many N_4GN_3 PAMs are inactive, and revealed no off-target sites with fewer than four mismatches in the mouse genome.

Figure 51 presents exemplary data showing that Nme1Cas9-mediated knockout of *Hpd* rescues the lethal phenotype in hereditary tyrosinemia Type I mice.

Figure 51A: Schematic diagram of the *Hpd* mouse gene. Red bars represent exons. Zoomed-in views show the protospacer sequences (red) for targeting exon 8 (sgHpd1) and exon 11 (sgHpd2). Nme1Cas9 PAM sequences are in green and double-stranded break locations are indicated (black arrowheads).

Figure 51B: Experimental design. Three groups of Hereditary Tyrosinemia Type I $Fah^{-/-}$ mice are injected with PBS or all-in-one AAV-sgRNA-hNme1Cas9 plasmids sgHpd1 or sgHpd2.

Figure 51C: Weight of mice hydrodynamically injected with PBS (green), AAV-sgRNA-hNme1Cas9 plasmid sgHpd1 targeting *Hpd* exon 8 (red) or sgHpd2-targeting *Hpd* exon 11 (blue) were monitored after NTBC withdrawal. Error bars represent three mice for PBS and sgHpd1 groups and two mice for the sgHpd2 group. Data are presented as mean ± SD.

Figure 51D: Stacked histogram showing a representative percentage distribution of indels at *Hpd* in liver of $Fah^{-/-}$ mice obtained by TIDE after hydrodynamic injection of PBS or sgHpd1 and sgHpd2 plasmids. Livers were harvested at the end of NTBC withdrawal (day 43).

Figure 52 presents exemplary data showing average indel efficiencies of the guides presented in Figure 51.

Figure 53 presents exemplary histological photomicrographs showing that liver damage is substantially less severe in the sgHpd1- and sgHpd2-treated mice compared to $Fah^{mut/mut}$ mice injected with PBS, as indicated by the smaller numbers of multinucleated hepatocytes compared to PBS-injected mice.

Figure 54 presents AAV-delivery of Nme1Cas9 for in vivo genome editing.

Figure 54A: Experimental outline of AAV8-sgRNA-hNme1Cas9 vector tail-vein injections to target *Pcsk9* (sgPcsk9) and *Rosa26* (sgRosa26) in C57Bl/6 mice. Mice were sacrificed at 4 (n = 1) or 50 days (n = 5) post injection and liver

tissues were harvested. Blood sera were collected at days 0, 25, and 50 post injection for cholesterol level measurement.

Figure 54B: Serum cholesterol levels. p values are calculated by unpaired t test.

Figure 54C: Stacked histogram showing a representative percentage distribution of indels at *Pcsk9* or *Rosa26* in livers of mice, as measured by targeted deep-sequencing analyses. Data are presented as mean ± SD from five mice per cohort.

Figure 54D: A representative anti-PCSK9 western blot using total protein collected from day 50 mouse liver homogenates. A total of 2 ng of recombinant mouse PCSK9 (r-PCSK9) was included as a mobility standard. The asterisk indicates a cross-reacting protein that is larger than the control recombinant protein.

Figure 55 presents exemplary data showing that mice injected with AAV8-sgRNA-hNme1Cas9 generate anti-Nme1Cas9 antibodies.

Figure 56 presents exemplary data showing GUIDE-seq genome-wide specificities of Nme1Cas9. Data are presented as mean ± SD.

Figure 56A: Number of GUIDE-seq reads for the on-target (OnT) and off-target (OT) sites.

Figure 56B: Targeted deep sequencing to measure the lesion rates at each of the OT sites in Hepa1–6 cells. The mismatches of each OT site with the OnT protospacers is highlighted (blue). Data are presented as mean ± SD from three biological replicates.

Figure 56C: Targeted deep sequencing to measure the lesion rates at each of the OT sites using genomic DNA obtained from mice injected with all-in-one AAV8-sgRNA-hNme1Cas9 sgPcsk9 and sgRosa26 and sacrificed at day 14 (D14) or day 50 (D50) post injection.

Figure 57 presents exemplary data for *Tyrosinase* (*Tyr*) gene editing ex vivo by Nme2Cas9 in mouse zygotes, as related to Figure 58.

Figure 57A: Two sites in *Tyr* gene, each with N_4CC PAMs, were tested for editing in Hepa1-6 cells. The sgTyr2 guide exhibited higher editing efficiency and was selected for further testing.

Figure 57B: Seven mice survived post-natal development, and each exhibited coat color phenotypes as well as on-target editing, as assayed by TIDE.

Figure 57C: Indel spectra from tail DNA of each mouse from Figure 57B, 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 58 presents exemplary data of ex vivo Nme2Cas9 genome editing using an all-in-one AAV delivery.

Figure 58A: 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:sgTyr for 5-6 hours, rinsed in M2, and cultured for a day before being transferred to the oviduct of pseudo-pregnant recipients.

Figure 58B: Albino (left) and chinchilla or variegated (middle) mice generated by 3 x 109 GCs, and chinchilla or variegated mice (right) generated by 3 x 108 GCs of zygotes with AAV6.Nme2Cas9:sgTyr.

Figure 58C: Summary of Nme2Cas9.sgTyr single-AAV ex vivo *Tyr* editing experiments at two AAV doses.

Figure 59 shows an alignment of Nme1Cas9 and Nme2Cas9 nucleotide sequences. Legend: Non-PID aa differences (turquoise shading); PID aa differences (yellow shading); active site residues (red letters).

Figure 60 shows an alignment of Nme1Cas9 and Nme3Cas9 nucleotide sequences. Legend: Non-PID aa differences (turquoise shading); PID aa differences (yellow shading); active site residues (red letters).

Figure 61 shows one embodiment of an Nme2Cas9 amino acid sequence. Legend: SV40 NLS (yellow shading); 3X-HA-Tag (green shading); cMyc-like NLS (turquoise shading); Linker (purple shading).

Figure 62 shows one embodiment of an Nme2Cas9 amino acid sequence. Legend: SV40 NLS (yellow shading); 3X-HA-Tag (green shading); Nucleoplasmin-like NLS (red shading); c-myc NLS (turquoise shading); Linker (purple shading).

Figure 63 shows one embodiment of a recombinant Nme2Cas9 (rNme2Cas9) amino acid sequence. Legend: SV40 NLS (yellow shading); Nucleoplasmin-like NLS (red shading); Linker (purple shading).

Figure 64 shows one embodiment of a all-in-one AAV-sgRNA-hNmeCas9 plasmid Nucleotide sequence. Legend: sgRNA scaffold (brown letters); GUIDE sequence (black letters); *U6 promoter* (blue letters); *U1a promoter* (purple letters): NLS NLS (green letters); hNmeCas9 (red letters); NLS 3X-HA and NLS BGH-pA (alternating green/black letters).

Detailed Description Of The Invention

The present invention is related to compositions and methods for gene therapy. Several approaches described herein utilize the *Neisseria meningitidis* Cas9 system that provides a hyperaccurate CRISPR gene editing platform. Furthermore, the invention incorporates improvements of this Cas9 system: for example, truncating the single guide RNA sequences, and the packing of -Nme1Cas9 or Nme2Cas9 with its guide RNA in an adeno-associated viral vector that is compatible for *in vivo* administration. Furthermore, Type II-C Cas9 orthologs have been identified that target protospacer adjacent motif sequences limited to between one – four required nucleotides.

I. *Neisseria meningitidis* Cas9 (Nme1Cas9)/CRISPR Gene Editing Accuracy

Previously, a hyper-accurate version of type II-C CRISPR-Cas9 systems called *Neisseria meningitidis* Cas9 (Nme1Cas9) was reported. In addition to being hyper-accurate, Nme1Cas9 is also smaller than the widely used *Streptococcus pyogenes* Cas9 (SpyCas9), allowing Nme1Cas9 to be delivered more readily via viral and messenger RNA (mRNA)-based methods. Genome editing with Nme1Cas9 typically has been accomplished using plasmid transfections. Zhang et al., "Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis" *Mol Cell* 50:488-503 (2013); Hou et al., "Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis" *Procd Natl Acad Sci U.S.A.* 110:15644-15649 (2013); Esvelt et al., "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing" *Nature Methods* 10:1116-1121 (2013); Zhang et al., "DNase H activity of Neisseria meningitidis Cas9" *Mol Cell* 60:242-255 (2015); Lee et al., "The Neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells"

Molecular Therapy 24:645-654 (2016); Pawluk et al., "Naturally occurring off-switches for CRISPR-Cas9" *Cell* 167:1829-1838 (2016); and Amrani et al., "Nme1Cas9 is an intrinsically high-fidelity genome editing platform" *biorxiv.org/content/early/2017/08/04/172650* (2017).

However, Nme1Cas9 viral, RNA- and ribonucleoproteins (RNP)-based delivery has not been extensively explored. RNA- and RNP-based delivery of Cas9 orthologs for genome engineering holds several advantages over other delivery methods. They not only result in faster editing since they bypass the expression issues related to DNA-based delivery of Cas9 and its sgRNA, but they also reduce off-target effects associated with Cas9-based editing. Reduced off-target activity results from finer control of the Cas9 RNA and RNP concentrations, and from relatively rapid Cas9 RNA and RNP degradation in cells. Prolonged presence of active Cas9 within the cell has been shown to be associated with higher off-target effects. Since Cas9 RNAs and RNPs are more rapidly degraded within cells, Cas9 delivered as RNA or RNP does not persist for long periods of time and consequently have reduced off-target effects.

Conventionally used full-length 145 nt Nme1Cas9 sgRNA includes a 48 nucleotide (nt) crRNA, a 4 nt linker, and a 93 nt tracrRNA. The crRNA region of the sgRNA is composed of a first 24 nt spacer sequence, and a second 24 nt repeat sequence that pairs with a 24 nt tracrRNA anti-repeat 5' region thereby forming a repeat:anti-repeat region. The remaining 69 nt tracrRNA region includes the Stem 1 region and Stem 2 region. Figure 1.

This full-length Nme1Cas9 sgRNA has been successfully used for genome editing using plasmid-based methods. Furthermore, *in vitro* transcribed Nme1Cas9 sgRNA can be complexed with purified Nme1Cas9 and used for genome editing in human cells. While genome editing of human cells has been successful with *in vitro* transcribed sgRNAs, the editing efficiency of an Nme1Cas9 RNP is reduced in harder-to-transfect human cell lines such as PLB985.

It has previously been shown that the editing efficiency of Cas9 RNPs is proportional to the chemical stability their sgRNAs. Although it is not necessary to understand the mechanism of an invention, it is believed that several cellular mechanisms are employed to rapidly degrade RNAs. For this reason, Cas9 sgRNAs are routinely modified by chemical means. Some of the chemical modifications that confer increased stability to sgRNA include, but are not limited to, ribose 2'-*O*-methylation and/or phosphorothioate linkages. While chemically modified RNAs are options for improved genome editing by Cas9 RNPs, their effectiveness is limited by the fact that chemical synthesis of RNAs becomes increasingly difficult and expensive as the length of

RNA increases. At 145 nt, Nme1Cas9 sgRNA synthesis is out of reach for routine genome editing applications that employ chemically synthesized sgRNAs.

II. Truncated Nme1Cas9 sgRNA Sequences

Due to the above identified limitation that a full-length 145 nt Nme1Cas9 sgRNA is too large for routine chemical synthesis of sgRNAs for genome editing, one embodiment of the present invention contemplates a truncated Nme1Cas9 sgRNA. Although it is not necessary to understand the mechanism of an invention, it is believed that a truncated Nme1Cas-sgRNA does not compromise the function of an Nme1Cas9 RNP. Furthermore, sgRNAs for Nme1Cas9 and Nme2Cas9 are identical and interchangeable (Figure 35B), so sgRNA truncations are equally applicable to both Nme1Cas9 and Nme2Cas9. Exemplary sequences of truncated sgRNAs and associated target sites are disclosed below, where variable sgRNA nts in guide regions are given as "N" residues. In the target sequences, the 24 nts recognized by the sgRNA guide region are underlined, and the protospacer adjacent motif (PAM) region is given in bold. Table 1.

Table 1: Exemplary Truncated sgRNA Sequences And Associated Genomic Targets

Description	Sequence
wt sgRNA	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCUUUCUCAUUUCGGAAACGAAAUGAGAACCGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCACGCUCUGCCCCUUAAAGCUUCUGCUUUAAGGGGCAUCGUUUA
sgRNA #1	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUAAAGCUUCUGCUUUAAGGGGCAUCGUUUA
sgRNA #2	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUUUCUAAGGGGCAUCGUUUA
sgRNA #3	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUUUCUAAGGGGCAU
sgRNA #4	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCUUCUGCAUCGUU
sgRNA #5	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCUUCUGCAUCGUUUA
sgRNA #6	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCUUCUGGGGCAUCGUU
sgRNA #7	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUUCUAGGGGCAUCGUU
sgRNA #8	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUCUGGGGCAUCGUU
sgRNA #9	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCCUUCUGGGCAUCGUU
sgRNA #10	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCUUCUGGGCAUCGUU
sgRNA #11	NNNNNNNNNNNNNNNNNNNNNNNNNNGUUGUAGCUCCCGAAACGUUGCUACAAUAAGGCCGUCUGAAAAGAUGUGCCGCAACGCUCUGCCUUCUGGCAUCGUU
N-TS7 Spacer (24 nt)	GAGGGAGAGAGGUGAGCGGAUGAA
N-TS7 Spacer (23 nt)	GGGAGAGAGGUGAGCGGAUGAA
N-TS27 Spacer (24 nt)	GUUCUCCAAGCCCUCGGACCUCGU
N-TS27 Spacer (23 nt)	GUCUCCAAGCCCUCGGACCUCGU
N-TS55 Spacer (24 nt)	GCUGGAUUACUGUGUGGUAGAGGG
N-TS55 Spacer (23 nt)	GUGGAUUACUGUGUGGUAGAGGG
N-TS7 Genomic Target Site	AGCTTGAGCAAAGGGAGAGAGGTGAGCGGATGAAGGGAGATTGGTGAGTATC
N-TS27 Genomic Target Site	CGCTTCGCGGCTTCTCCAAGCCCTCGGACCTCGTGGGCGTCTTCTCCTGCGT
N-TS55 Genomic Target Site	GAATTCACTAGCTGGATTACTGTGTGGTAGAGGGAGGTGATTAGCACCTGTG

As contemplated herein, a truncated Nme1Cas9 sgRNA would not only allow synthesis at a reasonable cost, but also facilitates use in virus-based delivery methods (e.g., for example adeno-associated viral delivery platforms) where the allowed length of DNA is limited. In one embodiment, the truncated sgRNA reduces off-target Nme1Cas9 editing effect. In one

embodiment, the truncated Nme1Cas9 sgRNA comprises at least one chemical modification that increases Nme1Cas9 editing efficiency.

As discussed above, the full length 145 nt sgRNA of Nme1Cas9 includes a guide region, a repeat:anti-repeat duplex region, a Stem 1 region and a Stem 2 region. Figure 1. However, because the length of the sgRNA is problematic for routine genomic editing, and it was highly desirable to develop a truncated sgRNA for Nme1Cas9. Currently, commercially available RNA synthesis methods require that RNA end product be not more than ~100 nt.

In one embodiment, the present invention contemplates an Nme1Cas9 sgRNA comprising a truncated repeat:anti-repeat duplex. In one embodiment, the present invention contemplates an Nme1Cas9 sgRNA comprising a truncated stem 2. Figure 2. Furthermore, it has previously been shown that a 5' variable guide crRNA region (e.g., spacer region) of Nme1Cas9 can also be truncated by a few nucleotides without loss of function. Amrani et al., "Nme1Cas9 is an intrinsically high-fidelity genome editing platform" *biorxiv.org/content/early/2017/08/04/172650* (2017); and Lee et al., "The Neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells" *Molecular Therapy* 24:645-654 (2016).

In one embodiment, the present invention contemplates a 100 nt Nme1Cas9-truncated sgRNA. Figure 3, Construct #11. This 100 nt Nme1Cas9 truncated-sgRNA Construct #11 was tested on three different human genomic sites by transient transfections in HEK293T cells, and at all three sites they support Nme1Cas9 function at the same level as, if not better than, the full-length Nme1Cas9 sgRNA. Figure 3, Bottom Panel. Moreover, sgRNA 11 and sgRNA 13 were also tested at several genomic target sites using RNP delivery and editing efficiency was similar or higher than the wt sgRNA. Figure 4. The synthetic version of construct #11 was also tested in PLB985 cells resulting in higher editing efficiency relative to in vitro transcribed wt sgRNA. Figure 5.

III. Associated-Adenovirus CRISPR Delivery Platforms

Compared to transcription activator-like effector nucleases (TALENs) and Zinc-finger nucleases (ZFNs), Cas9s are distinguished by their flexibility and versatility. Komor et al., "CRISPR-based technologies for the manipulation of eukaryotic genomes" *Cell* 2017;168:20–36 . Such characteristics make them ideal for driving the field of genome engineering forward. Over the past few years, CRISPR-Cas9 has been used to enhance products in agriculture, food,

and industry, in addition to the promising applications in gene therapy and personalized medicine. Barrangou et al., "Applications of CRISPR technologies in research and beyond" *Nat Biotechnol.* 2016;34:933–41. Despite the diversity of Class 2 CRISPR systems that have been described, only a handful of them have been developed and validated for genome editing in vivo. As shown herein, NmeCas9 is a compact, high-fidelity Cas9 that can be considered for future in vivo genome editing applications using all-in-one rAAV. NmeCas9's unique PAM enables editing at additional targets that are inaccessible to the other two compact, all-in-one rAAV-validated orthologs (SauCas9 and CjeCas9).

Genome editing using a bacterial CRISPR system has opened a new avenue for human gene therapy. Named for Clustered Regularly Interspaced Short Palindromic Repeats that capture snippets of invasive nucleic acids in bacteria, the CRISPR complex comprises a guide RNA (e.g., sgRNA) that directs a nuclease Cas9 (CRISPR-associated protein 9) to cleave complementary double-stranded DNA. Non-homologous repair of a Cas9-induced DNA break leads to small insertions or deletions (indels) that inactivate target genes, but breaks can also be repaired by homologous DNA templates resulting in gene replacement. Nelson et al., "In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy" *Science* 351: 403-407 (2016); and Ran et al., "In vivo genome editing using Staphylococcus aureus Cas9" *Nature* 520:186-191 (2015); and Yin et al., "Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype" *Nature Biotechnology* 32:551-553 (2014).

The current and widely-used Type II-A *Streptococcus pyogenes* (Spy) Cas9 as a flexible genome-editing tool demonstrates several disadvantages: i) inefficient delivery; ii) off-target cleavage; and iii) unregulated activity. These disadvantages strictly limit SpyCas9 as a potential gene therapy tool. As discussed herein a highly accurate and precise Nme1Cas9 or Nme2Cas9 complex can overcome these SpyCas9 limitations.

Nme1Cas9 and Nme2Cas9 have been shown herein to be an efficient genome-editing platform in mammalian cells and, as a smaller protein than SpyCas9, it is easier to engineer viral vectors for in vivo delivery. Furthermore, Nme1Cas9 and Nme2Cas9 have significantly lower off-target editing than SpyCas9 and anti-CRISPR proteins have been identified that allow control of Nme1Cas9 and Nme2Cas9 activity. Esvelt et al., "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing" *Nature Methods* 10:1116-1121 (2013); Amrani et al., "Nme1Cas9 is an intrinsically high-fidelity genome editing platform" *biorxiv.org/content/early/2017/08/04/*

172650 (2017); Lee et al., 'The *Neisseria meningitidis* CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells" *Molecular Therapy* 24:645-654 (2016); Hou et al., "'Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*" *Procd Natl Acad Sci USA* 110:15644-15649 (2013); and Pawluk et al., "Naturally Occurring Off-Switches for CRISPR-Cas9" *Cell* 167:1829-38 e9 (2016); and Figure 21.

Adeno-Associated Virus (AAV) has been demonstrated as a delivery shuttle with minimal pathogenicity in pre-clinical and clinical settings, but it has a limited packaging capacity. Nme1Cas9, encoded by a ~3.3kb open reading frame, and its guide RNAs are within the packaging limit of AAV. Nme2Cas9 has similar advantages. Unlike SpyCas9, which requires delivery by separate vectors for the sgRNA and Cas9, Nme1Cas9, Nme2Cas9 and their sgRNA are small enough to be delivered with a single AAV vector.

Other Cas9 orthologs have been successfully delivered in vivo by AAV, such as *Campylobacter jejuni* Cas9 (CjeCas9) and *Staphylococcus aureus* (SauCas9). Kim et al., "In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*" *Nat Commun* 8:14500 (2017); and Ran et al., "In vivo genome editing using *Staphylococcus aureus* Cas9" *Nature* 520:186-191 (2015). Nme1Cas9 is usually associated with an N_4GATT PAM, which is unlike the CjeCas9 PAM (e.g., N_4RYAC), or the SauCas9 PAM (e.g., NNGRRT) (R = purine (A or G), Y = pyrimidine (C or T)).

Nme1Cas9 has been successfully delivered as a ribonucleoprotein (RNP) complex in human cells. Figure 2 and Figure 3. Further, the data presented herein show that an Nme1Cas9 nucleic acid sequence can be expressed in vivo in mice to target genes using an all-in-one sgRNA-Nme1Cas9-AAV vector subsequent to a tail vein injection.

The data presented herein demonstrates a targeting of a mouse Proprotein Convertase Subtilisin/Kexin type 9 (*Pcsk9*) gene. PCSK9 functions as an antagonist to the low-density lipoprotein (LDL) receptor and limits LDL cholesterol uptake. Detection of reduced cholesterol levels in the serum can thereby provide a direct functional readout of efficient Nme1Cas9 editing using a *PCSK9* - directed Cas9 platform.

In one embodiment, the present invention contemplates an adeno-associated viral vector comprising an Nme1Cas9-sgRNA complex or an Nme2Cas9-sgRNA complex. Although it is not necessary to understand the mechanism of an invention, it is believed that an

AAV/Nme1Cas9-sgRNA complex or an AAV/Nme2Cas9-sgRNA complex are compatible with an in vivo delivery route in order to provide gene editing.

In one embodiment, the present invention contemplates an sgRNA-Nme1Cas9-AAV vector comprising an sgRNA sequence, an RNA Polymerase III U6 promoter sequence, a human codon-optimized Nme1Cas9 sequence, and an RNA Polymerase II U1a promoter sequence. Figure 6. U1a is a ubiquitous promoter allowing versatile expression of Cas9 in various tissues of interest. Specific genes to be edited can be targeted by inserting a spacer sequence matching a target gene into an sgRNA cassette using conventional restriction sites (e.g., Sap1). Representative sequences of the various elements of the sgRNA-Nme1Cas9-AAV are shown by colored annotations. Figures 7 and 8.

Editing efficiencies of several target sites using a *Pcsk9*-sgRNA-Nme1Cas9-AAV plasmid and a *Rosa26*-sgRNA-Nme1Cas9-AAV plasmid were estimated by an T7E1 assay following transient transfection into mouse Hepa1-6 hepatoma cells. Figure 9. Representative target site sequences within a *Pcsk9* gene and a *Rosa26* gene complementary with a *Pcsk9*-sgRNA-Nme1Cas9-AAV plasmid and a *Rosa26*-sgRNA-Nme1Cas9-AAV plasmid are shown by colored annotations. Figure 10.

The plasmid design was validated in vivo with mice by hydrodynamic injection of 30 μg of endotoxin-free sgRNA-Nme1Cas9-AAV plasmid targeting *Pcsk9* via tail-vein. Significant gene editing was detected in mouse liver 10 days after injection as measured by Tracking of Indels by DEcomposition (TIDE), a sequencing-based method of evaluating indel efficiencies. Figure 11.

The plasmid backbones targeting a *Pcsk9* gene and a *Rosa26* gene were packaged in hepatocyte-specific AAV8 serotype, and a dose of $4x10^{10}$ genomic copies (gc) per mouse was injected via tail-vein. Preliminary data show indel values from mice sacrificed at 14 days post-injection at a significant indel level in liver *Pcsk9* and *Rosa26* genes. Figure 12A. Deep-sequencing data has also been collected at day 50 post-injection.

The three mice groups were sacrificed at day 50 post-injection, and liver gDNA was used to measure the indel values at *Pcsk9* and *Rosa26* using TIDE. Figure 12B. Deep-sequencing analyses has also been performed to record accurate measurements of indel values.

PCSK9 protein "knock-down" may lead to significant lowering of cholesterol levels in mice. Serum cholesterol level was measured by Infinity™ colorimetric endpoint assay (Thermo-

Scientific) in 3 mice groups injected with vectors targeting a *Pcsk9* gene, a *Rosa26* gene and a PBS control group. Results suggest that Nme1Cas9-induced indel formation has led to the interruption of the normal reading frame of the *Pcsk9* gene, as showed by significantly reduced values of serum cholesterol at 25 and 50 days post-injection. Figure 13. Western blot assay has also been performed to measure the level of PCSK9 protein in mice liver at day 50.

A genome-wide unbiased identification of double strand breaks (DSBs) enabled by a sequencing assay (e.g., GUIDE-Seq®, Illumina) searched for off-target editing sites subsequent to injection of vectors targeting a *Pcsk9* gene and a *Rosa26* gene. The data revealed four (4) potential off-target sites for *Pcsk9* and six (6) potential off-target sites for *Rosa26*. Figures 14A and 14B.

A targeted TIDE analyses revealed on-target genome editing in cells and in the mice at day 14 subsequent to injection of AAV vectors targeting a *Pcsk9* gene and a *Rosa26* gene. Figure 15. Deep-sequencing analyses for off-target cleavage at these sites has also been performed at 50 days post-injection.

A hematoxylin and eosin stain assay did not show signs of massive immune cell infiltration in the liver sections of mice sacrificed at day 14 subsequent to injection of vectors targeting a *Pcsk9* gene and a *Rosa26* gene. Figure 16. Specific immune-response assays will be performed at 50 day post-injection.

In one embodiment, the present invention contemplates a method for therapeutic *in vivo* genome editing by all-in-one AAV delivery of an Nme2Cas9. Although it is not necessary to understand the mechanism of an invention it is believed that the compactness, small PAM and high fidelity make Nme2Cas9 an ideal tool for *in vivo* genome editing using AAV. To this end, Nme2Cas9 was cloned with its cognate sgRNA and their respective promoters into a single AAV vector backbone. Figure 44A; top.. This all-in-one AAV.sgRNA.Nme2Cas9 was packaged in a hepatocyte-selective AAV8 capsid. Two genes were targeted: i) *Rosa26*, a commonly used locus as a negative control; and ii) the *Proprotein convertase subtilisin/kexin type 9* (*Pcsk9*), a major regulator of circulating cholesterol homeostasis. Studies have shown that knocking out *Pcsk9* using Cas9 results in reduced cholesterol levels (Ran et al).

Two groups of mice ($n=5$) were injected with packaged AAV8.sgNA.Nme2Cas9 targeting either *Pcsk9* or *Rosa26*. Serum was collected at 0, 14 and 28 days post vector injection for cholesterol level measurement. Mice were sacrificed at 28 days post-injection

and liver tissues were harvested. (Figure 44A, bottom. A deep sequencing analysis showed significantly high level of indels at *Pcsk9* and *Rosa26*. Figure 44B.. These indel values were accompanied by significant reduction in blood cholesterol level in mice injected with *sgPcsk9* after 14 and 28 days; where mice injected with *sgRosa26* maintained normal level of cholesterol throughout the study. Figure 44C. An H&E analyses showed no signs of toxicity or tissue damage at both groups after Nme2Cas9 expression. Figure 44D.. These data validate that Nme2Cas9 is highly functional *in* vivo, and it can be readily delivered by the favorable all-in-one AAV platform.

In one embodiment, the present invention contemplates a minimized AAV.hNmeCas9 construct. See, Figure 44A. As discussed above, the present invention contemplates an engineered all-in-one AAV.sgRNA.hNme1Cas9 construct, which is packaged in AAV8 virions that successfully edited *Pcsk9* and *Rosa26* genes in mice liver.

In one embodiment, the present invention contemplates an AAV8 backbone comprising an Nme2Cas9 cassette. Similar to Nme1Cas9, Nme2Cas9 also showed robust editing at *Pcsk9* and *Rosa26* in mice (*infra*). The data presented herein shows that in vivo administration of AAV8-NmeCas9 to mice is accompanied by significant reduction in level of circulating cholesterol after 28 days post vector injection.

In order to increase the utility of this all-in-one AAV platform, various truncations were introduced to minimize the size of the cargo to make a space for additional features in the AAV capsid, such as dual sgRNAs or donor DNA segment.

In order to minimize the cargo of the all-in-one AAV backbone, the extra features (3x HA tags and 2x NLS sequences) were systematically removed without compromising the nuclease activity of the Cas9. Nme1Cas9, using the traffic light reporter (TLR) system, show that this minimized all-in-one AAV.sgRNA.hNme1Cas9 (4.468 kb) is as potent as the previous longer version with 4 NLS sequences. See, Figure 45. Truncated sgRNAs were constructed to free more space using a new sgRNA12, which is similar to an sgRNA11 version, but with UA added at the 3' end. See, Figure 46..

Previously, it has been reported that a short polyA sequence may be useful for Cas9 constructs. Platt *et. al.* (2015). In one embodiment, the present invention contemplates an AAV-Nme2Cas9 construct comprising a BGH polyA. See, Figure 47. Although it is not necessary to

understand the mechanism of an invention, it is believed that this polyA sequence further reduces the size of the all-in-one AAV backbone.

It is further believed that this minimized (4.4 kb) all-in-one AAV backbone increases the utility of Nme1Cas9 and Nme2Cas9 by including another sgRNA for dual genes knockout or DNA fragment excision. See, Figure 48, top. This configuration also provides free space in the AAV capsid to include a donor template (~ 600 base pairs) for homology-directed repair application. See, Figure 48, bottom. In some embodiments, dual sgRNA AAV constructs are packaged within a single AAV vector.

The relatively compact Nme1Cas9 is active in genome editing in a range of cell types. To exploit the small size of this Cas9 ortholog, an all-in-one AAV construct was generated with human-codon-optimized Nme1Cas9 under the expression of the mouse U1a promoter and with its sgRNA driven by the U6 promoter. See, Figure 49A. Two sites in the mouse genome were selected initially to test the nuclease activity of Nme1Cas9 in vivo: the *Rosa26* "safe-harbor" gene (targeted by sgRosa26); and the proprotein convertase subtilisin/kexin type 9 (*Pcsk9*) gene (targeted by sgPcsk9), a common therapeutic target for lowering circulating cholesterol and reducing the risk of cardiovascular disease. Figure 49B. Genome-wide off-target predictions for these guides were determined computationally using the Bioconductor package CRISPRseek 1.9.1 with N_4GN_3 PAMs and up to six mismatches. Zhu et al., "CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genomeediting systems" *PLoS One* 2014;9:e108424. Many N_4GN_3 PAMS are inactive, so these search parameters are nearly certain to cast a wider net than the true off-target profile. Despite the expansive nature of the search, an analyses revealed no off-target sites with fewer than four mismatches in the mouse genome. See, Figure 50. On-target editing efficiencies at these target sites were evaluated in mouse Hepa1–6 hepatoma cells by plasmid transfections and indel quantification was performed by sequence trace decomposition using the Tracking of Indels by Decomposition (TIDE) web tool. Brinkman et al., "Easy quantitative assessment of genome editing by sequence trace decomposition" *Nucleic Acids Res.* 2014;42:e168. The data show > 25% indel values for the selected guides, the majority of which were deletions. See, Figure 49C.

To evaluate the preliminary efficacy of the constructed all-in-one AAV-sgRNA-hNme1Cas9 vector, endotoxin-free sgPcsk9 plasmid was hydrodynamically administered into the C57Bl/6 mice via tail-vein injection. This method can deliver plasmid DNA to ~ 40% of

hepatocytes for transient expression. Liu et al., "Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA" *Gene Ther.* 1999;6:1258–66. Indel analyses by TIDE using DNA extracted from liver tissues revealed 5–9% indels 10 days after vector administration, comparable to the editing efficiencies obtained with analogous tests of SpyCas9. See, Figure 49D; and Xue et al., "CRISPR-mediated direct mutation of cancer genes in the mouse liver" *Nature* 2014;514:380–4. These results suggest that Nme1Cas9 is capable of editing liver cells in vivo.

Hereditary Tyrosinemia type I (HT-I) is a fatal genetic disease caused by autosomal recessive mutations in the *Fah* gene, which codes for the fumarylacetoacetate hydroxylase (FAH) enzyme. Patients with diminished FAH have a disrupted tyrosine catabolic pathway, have a disrupted tyrosine catabolic pathway, leading to the accumulation of toxic fumarylacetoacetate and succinyl acetoacetate, causing liver and kidney damage. Grompe M., "The pathophysiology and treatment of hereditary tyrosinemia type 1" *Semin Liver Dis.* 2001;21:563–71. Over the past two decades, the disease has been controlled by 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC), which inhibits 4-hydroxyphenylpyruvate dioxygenase upstream in the tyrosine degradation pathway, thus preventing the accumulation of the toxic metabolites. Lindstedt et al., "Treatment of hereditary tyrosinaemia type I by inhibition of 4-hydroxyphenylpyruvate Dioxygenase" *Lancet* 1992;340:813–7. However, this treatment requires lifelong management of diet and medication and may eventually require liver transplantation. Das, AM., "Clinical utility of nitisinone for the treatment of hereditary tyrosinemia type-1 (HT-1)" *Appl Clin Genet.* 2017;10:43–8.

Several gene therapy strategies have been tested to correct a defective *Fah* gene using site-directed mutagenesis or homology-directed repair by CRISPR-Cas9. Paulk et al., "Adenoassociated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo" *Hepatology* 2010;51:1200–8; Yin et al., "Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo" *Nat Biotechnol.* 2016;34:328–33; and Yin et al., "Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype" *Nat Biotechnol.* 2014;32:551–3. It has been reported that successful modification of only 1/10,000 of hepatocytes in the liver is sufficient to rescue the phenotypes of $Fah^{mut/mut}$ mice. Recently, a metabolic pathway reprogramming approach has been suggested in which the

function of the hydroxyphenylpyruvate dioxygenase (HPD) enzyme was disrupted by the deletion of exons 3 and 4 of the *Hpd* gene in the liver. Pankowicz et al., "Reprogramming metabolic pathways in vivo with CRISPR/Cas9 genome editing to treat hereditary tyrosinaemia" *Nat Commun.* 2016;7:12642. This provides a context in which to test the efficacy of Nme1Cas9 editing, for example, by targeting *Hpd* and assessing rescue of the disease phenotype in *Fah* mutant mice. Grompe et al., "Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice" *Genes Dev.* 1993;7:2298–307. For this purpose, two target sites (one each in exon 8 [sgHpd1] and exon 11 [sgHpd2]) were screened and identified within the open reading frame of *Hpd*. See, Figure 51A. These guides (e.g., sgRNAs) facilitated Nme1Cas9-induced average indel efficiencies of 10.8% and 9.1%, respectively, by plasmid transfections in Hepa1–6 cells. Figure 52.

Three groups of mice were treated by hydrodynamic injection with either phosphate-buffered saline (PBS) or with one of the two sgHpd1 and sgHpd2 all-in-one AAV-sgRNA-hNme1Cas9 plasmids. One mouse in the sgHpd1 group and two in the sgHpd2 group were excluded from the follow-up study due to failed tail-vein injections. Mice were taken off NTBC-containing water seven days after injections and their weight was monitored for 43 days post injection. See, Figure 51B. Mice injected with PBS suffered severe weight loss (a hallmark of HT-I) and were sacrificed after losing 20% of their body weight. Overall, all sgHpd1 and sgHpd2 mice successfully maintained their body weight for 43 days overall and for at least 21 days without NTBC. See, Figure 51C.

NTBC treatment had to be resumed for 2–3 days for two mice that received sgHpd1 and one that received sgHpd2 to allow them to regain body weight during the third week after plasmid injection, perhaps due to low initial editing efficiencies, liver injury due to hydrodynamic injection, or both. Conversely, all other sgHpd1 and sgHpd2 treated mice achieved indels with frequencies in the range of 35–60%. See, Figure 51D. This level of gene inactivation likely reflects not only the initial editing events but also the competitive expansion of edited cell lineages (after NTBC withdrawal) at the expense of their unedited counterparts.. Liver histology revealed that liver damage is substantially less severe in the sgHpd1- and sgHpd2-treated mice compared to $Fah^{mut/mut}$ mice injected with PBS, as indicated by the smaller numbers of multinucleated hepatocytes compared to PBS-injected mice. See, Figure 53.

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. Editing was obtained previously with a dual-AAV system in which SpyCas9 and its sgRNA were delivered in separate vectors. Yoon et al., "Streamlined ex vivo and in vivo genome editing in mouse embryos using recombinant adeno-associated viruses" *Nat. Commun.* 9:412 (2018). To test whether Nme2Cas9 could enable accurate and efficient editing in mouse zygotes with an all-in-one AAV delivery system, the tyrosinase gene (*Tyr*) was targeted, where a bi-allelic inactivation of which disrupts melanin production, resulting in albino pups. Yokoyama et al., "Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice" *Nucleic Acids Res.* 18:7293-7298 (1990).

An efficient *Tyr* sgRNA (which cleaves the *Tyr* locus only 17 bp from the site of the classic albino mutation) was validated in Hepa1-6 cells by transient transfections. See, Figure 57. Next, C57BL/6NJ zygotes were incubated for 5-6 hours in culture medium containing 3 x 10^9 or 3 x 10^8 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 58A. Coat color analysis of pups revealed mice that were albino, light grey (suggesting a hypomorphic allele of *Tyr*), or that had variegated coat color composed of albino and light grey spots but lacking black pigmentation. See, Figures 58B & 58C. These results suggest a high frequency of biallelic mutations since the presence of a single wild-type *Tyr* allele should render black pigmentation. A total of five pups (10%) were born from the 3 x 10^9 GCs experiment. All of them carried indels; phenotypically, two were albino, one was light grey, and two had variegated pigmentation, indicating mosaicism. From the 3 x 10^8 GCs experiment, four (4) pups (14%) were obtained, two of which died at birth, preventing coat color or genome analysis. Coat color analysis of the remaining two pups revealed one light grey and one mosaic pup. These results indicate that single-AAV delivery of Nme2Cas9 and its sgRNA 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 a TIDE analysis was performed. The data showed that all mice had high levels of on-target editing by Nme2Cas9, varying from 84% to 100%. See,

Figures 57B and 5C. Most lesions in albino mouse 9-1 were either a 1- or a 4-bp deletion, suggesting either mosaicism or trans-heterozygosity. Albino mouse 9-2 exhibited a uniform 2-bp deletion. See, Figure 58C. Analysis of tail DNA from light grey mice revealed the presence of in-frame mutations that are potentially a cause of the light grey coat color. The limited mutational complexity suggests that editing occurred early during embryonic development in these mice. One female (mouse 9-2) was mated with a classical albino male, and all six of the resulting pups were albino, demonstrating that mutations generated by zygotic all-in-one AAV delivery of Nme2Cas9 + sgRNA can be transmitted through the germline. 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.

Patients with mutations in the *Hpd* gene are considered to have Type III Tyrosinemia and exhibit high level of tyrosine in blood, but otherwise appear to be largely asymptomatic. Szymanska et al., "Tyrosinemia type III in an asymptomatic girl. Mol Genet Metab Rep. 2015;5:48–50; and Nakamura et al., "Animal models of tyrosinemia" *J Nutr.* 2007;137:1556S–60S. HPD acts upstream of FAH in the tyrosine catabolism pathway and *Hpd* disruption ameliorates HT-I symptoms by preventing the toxic metabolite build-up that results from loss of FAH. Structural analyses of HPD reveal that the catalytic domain of the HPD enzyme is located at the C-terminus of the enzyme and is encoded by exon 13 and 14. Huang et al., "The different catalytic roles of the metal-binding ligands in human 4-hydroxyphenylpyruvate dioxygenase" *Biochem J.* 2016;473:1179–89. Thus, frameshift-inducing indels upstream of exon 13 should render the enzyme inactive. This context was used to demonstrate that *Hpd* inactivation by hydrodynamic injection of Nme1Cas9 plasmid is a viable approach to rescue HT-I mice. Nme1Cas9 can edit sites carrying several different PAMs (N_4GATT [consensus], N_4GCTT, N_4GTTT, N_4GACT, N_4GATA, N_4GTCT, and N_4GACA). *Hpd* editing experiments confirmed one of the variant PAMs in vivo with the sgHpd2 guide, which targets a site with a N_4GACT PAM.

Although plasmid hydrodynamic injections can generate indels, therapeutic development may require less invasive delivery strategies, such as by using an rAAV. To this end, all-in-one AAV-sgRNA-hNme1Cas9 plasmids were packaged in hepatocyte-tropic AAV8 capsids to target *Pcsk9* (sgPcsk9) and *Rosa26* (sgRosa26). See, Figure 49B; Gao et al., "Novel adenoassociated

viruses from rhesus monkeys as vectors for human gene therapy" *Proc Natl Acad Sci U S A* 2002;99:11854–9; and Nakai et al., "Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice" *J Virol.* 2005;79:214–24. *Pcsk9* and *Rosa26* were used in part to enable Nme1Cas9 AAV delivery to be benchmarked with that of other Cas9 orthologs delivered similarly and targeted to the same loci. Ran et al., "In vivo genome editing using Staphylococcus aureus Cas9" *Nature* 2015;520:186–91. Vectors were administered into C57BL/6 mice via tail vein. See, Figure 54A. Cholesterol levels were monitored in the serum and measured PCSK9 protein and indel frequencies in the liver tissues 25 and 50 days post injection.

Using a colorimetric endpoint assay, it was determined that the circulating serum cholesterol level in the mice administered Nme1Cas9/sgPcsk9 decreased significantly ($p < 0.001$) compared to the PBS and Nme1Cas9/sgRosa26 mice at 25 and 50 days post injection. See, Figure 54B. Targeted deep-sequencing analyses at *Pcsk9* and *Rosa26* target sites revealed very efficient indels of 35% and 55%, respectively, at 50 days post vector administration. Figure 54C. Additionally, one mouse of each group was euthanized at 14 days post injection and revealed on-target indel efficiencies of 37% and 46% at *Pcsk9* and *Rosa26*, respectively. As expected, PCSK9 protein levels in the livers of Nme1Cas9/sgPcsk9 treated mice were substantially reduced compared to the mice injected with PBS and Nme1Cas9/sgRosa26. See, Figure 54D. The efficient editing, PCSK9 reduction, and diminished serum cholesterol indicate the successful delivery and activity of Nme1Cas9 at the *Pcsk9* locus.

SpyCas9 delivered by viral vectors is known to elicit host immune responses. Chew et al., "A multifunctional AAV-CRISPR-Cas9 and its host response" *Nat Methods* 2016;13:868–74; and Wang et al., "Adenovirus-mediated somatic genome editing of Pten by CRISPR/Cas9 in mouse liver in spite of Cas9-specific immune responses" *Hum Gene Ther.* 2015;26:432–42. To investigate if the mice injected with AAV8-sgRNA-hNme1Cas9 generate anti-Nme1Cas9 antibodies, sera was used from the treated animals to perform IgG1 ELISA. These results show that Nme1Cas9 elicits a humoral response in these animals. See, Figure 55. Despite the presence of an immune response, Nme1Cas9 delivered by rAAV is highly functional in vivo, with no apparent signs of abnormalities or liver damage. See, Figure 16.

A significant concern in therapeutic CRISPR/Cas9 genome editing is the possibility of activity at off-target edits. For example, it has been found that wild-type Nme1Cas9 is a

naturally high-accuracy genome editing platform in cultured mammalian cells. Lee et al., "The Neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells" *Mol Ther*. 2016;24:645–54. To determine if Nme1Cas9 maintains its minimal off-targeting profile in mouse cells and in vivo, off-target sites were screened in the mouse genome using genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq). Tsai et al., "Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases" *Nat Rev Genet.* 2016;17:300–12. Hepa1–6 cells were transfected with sgPcsk9, sgRosa26, sgHpd1, and sgHpd2 all-in-one AAV-sgRNA-hNme1Cas9 plasmids and the resulting genomic DNA was subjected to GUIDE-seq analysis. Consistent with observations in human cells (data not shown), GUIDE-seq revealed very few off-target (OT) sites in the mouse genome. Four potential OT sites were identified for sgPcsk9 and another six for sgRosa26. Off-target edits with sgHpd1 and sgHpd2 were not detected. See, Figure 56A. These data further validate that Nme1Cas9 is intrinsically hyper-accurate.

Several of the putative OT sites for sgPcsk9 and sgRosa26 lack the Nme1Cas9 PAM preferences (i.e., N_4GATT, N_4GCTT, N_4GTTT, N_4GACT, N_4GATA, N_4GTCT, and N_4GACA). See, Figure 56B. To validate these OT sites, targeted deep sequencing was performed using genomic DNA from Hepa1–6 cells. By this more sensitive readout, indels were undetectable above background at all these OT sites except OT1 of *Pcsk9*, which had an indel frequency < 2%. See, Figure 56B. To validate Nme1Cas9's high fidelity in vivo, indel formation was measured at these OT sites in liver genomic DNA from the AAV8-Nme1Cas9-treated, sgPcsk9-targeted, and sgRosa26-targeted mice. Little or no detectable off-target editing was found in mice liver sacrificed at 14 days at all sites except sgPcsk9 OT1, which exhibited < 2% lesion efficiency. More importantly, this level of OT editing stayed below < 2% even after 50 days and also remained either undetectable or very low for all other candidate OT sites. These results suggested that extended (50 days) expression of Nme1Cas9 in vivo does not compromise its targeting fidelity. See, Figure 56C.

To achieve targeted delivery of Nme1Cas9 to various tissues in vivo, rAAV vectors are a promising delivery platform due to the compact size of Nme1Cas9 transgene, which allows the delivery of Nme1Cas9 and its guide in an all-in-one format. The data presented herein validates this approach for the targeting of *Pcsk9* and *Rosa26* genes in adult mice, with efficient editing observed even at 14 days post injection. Nme1Cas9 is intrinsically accurate, even without the

extensive engineering that was required to reduce off-targeting by SpyCas9. Lee et al., "The Neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells" *Mol Ther*. 2016;24:645–54; Bolukbasi et al., "Creating and evaluating accurate CRISPRCas9 scalpels for genomic surgery" *Nat Methods* 2016;13:41–50; Tsai et al., "Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases" *Nat Rev Genet.* 2016;17:300–12; and Tycko et al., "Methods for optimizing CRISPR-Cas9 genome editing specificity" *Mol Cell*. 2016;63:355–70.

Side-by-side comparisons of Nme1Cas9 OT editing were performed in cultured cells and in vivo by targeted deep sequencing and found that off-targeting is minimal in both settings. Editing at the sgPcsk9 OT1 site (within an unannotated locus) was the highest detectable at ~ 2%.

IV. Small Cas9 Orthologs With Cytosine-Rich PAMs

As noted above, CRISPR systems may be classified into at least six (6) different types. Generally, Type II systems are categorized by the presence of a Cas9 nuclease protein. For example, a Cas9 nuclease protein is believed to be an RNA-guided nuclease that can be repurposed as a genome editing platform in almost all organisms, including humans. Reports have indicated that Cas9 genome editing has been used in medicine, agriculture, human gene therapy and many other applications.

Generally, targeting of a specific gene locus in the human genome may be accomplished by a Cas9 nuclease protein bound to a single guide RNA (sgRNA) that targets the locus via an interaction with a specific nucleic acid sequence (e.g., for example, a protospacer adjacent motif; PAM). sgRNA's usually comprise a 20-24 nucleotide segment that is complementary to a target nucleic acid sequence followed by a constant region that interacts (e.g., for example, binds) with the Cas9 protein. For the Cas9 nuclease protein to perform genome editing, the Cas9:sgRNA complex first recognizes a protospacer adjacent motif (PAM) sequence that is normally found downstream of the target site sequence. Although it is not necessary to understand the mechanism of an invention, it is believed that each Cas9 nuclease protein has affinity for a particular PAM (i.e., mediated by a protospacer adjacent motif recognition domain). In the absence of the PAM recognition domain binding to a downstream PAM target nucleic acid sequence double-stranded DNA (dsDNA) cannot be cleaved by the Cas9 nuclease.

Reports suggest that only a handful of Cas9 orthologs have been validated for human genome editing. Three of the reported CRISPR-Cas9 types include II-A, II-B and II-C. Type II-A Cas9 (e.g., *Streptococcus pyogenes* (SpyCas9)), is the most commonly used Cas9 to date. However, SpyCas9 (and most other type II-A orthologs) possesses several characteristics that may make it unsuitable for certain applications. First, SpyCas9 is relatively large, making this Cas9 unsuitable for efficient packaging into viral vectors. Second, SpyCas9 has a high rate of off-target activity (i.e. it cleaves DNA at unintended loci in the human genome), although higher-specificity variants have been engineered. Finally, SpyCas9's PAM (e.g., NGG) has limited use in some sites in the human genome, or for applications where a specific nucleotide is to be recognized during editing. To overcome these shortcomings, several groups have repurposed other Cas9 orthologs to function in humans and other organisms. As discussed above, type II-C Cas9 orthologs (e.g., Nme1Cas9) are small enough for all-in-one viral packaging (e.g., adeno-associated virus (AAV) vectors] that results in higher fidelity activity in mammalian cells. However, wild type Cas9 II-C PAMs are usually approximately four (4) nucleotides in length as opposed to an SpyCas9 PAM that is usually two (2) nucleotides in length. This additional PAM length can limit the number of loci that can be targeted by a wild type Cas9 II-C PAM. This creates a need in the art for the identification of more Cas9 orthologs for genome editing.

While there are thousands of Cas9 orthologs in the NCBI database to choose from, an empirical process is required to develop small type II-C Cas9 orthologs with less restrictive PAMs that provide improved functionality in mammalian cells. In one embodiment, the present invention contemplates an improved type II-C Cas9 ortholog that enables precise genome editing with a broader range of target sites. In one embodiment, the improved type II-C Cas9 ortholog has a compact size capable of efficient viral delivery. In one embodiment, the improved type II-C Cas9 ortholog includes, but is not limited to, *Haemophilus parainfluenzae* (HpaCas9), *Simonsiella muelleri* (SmuCas9) and *Neisseria meningitidis strain De10444* (Nme2Cas9).

A. Short PAMs Associated With Type II-C Cas9 Orthologs

The data presented herein shows the characterization of short PAM targets for several type II-C Cas9 orthologs. Figure 17. For example, type II-C Cas9 orthologs may interact with short PAMs comprising between one – four required nucleotides. Although it is not necessary to understand the mechanism of an invention, it is believed that these short C-rich PAMs provide

improved Cas9 genome editing of target sites previously not accessible even by the more compact Cas9 orthologs (e.g., Nme1Cas9). In one embodiment, an Nme2Cas9 PAM has a sequence of NNNNCc, wherein "c" is the only a partial preference. In one embodiment, an SmuCas9 PAM has a sequence of NNNNCT. Figure 18.

It is currently believed that no Cas9 orthologs with short C-rich PAMs have been validated for genome editing and that Nme2Cas9 is particularly compelling as a potential candidate for highly efficient gene editing activity in human cells. In one embodiment, the present invention contemplates an Nme2Cas9 nuclease bound to a wild type Nme1Cas9 sgRNA (e.g., *Neisseria meningitidis 8013* Cas9; previously referred to as NmeCas9). Nme1Cas9 has been previously described. Sontheimer et al., "RNA-Directed DNA Cleavage and Gene Editing by Cas9 Enzyme From *Neisseria Meningitidis*" United States Patent Application Publication Number 2014/0349,405 (herein incorporated by reference). Although Nme1Cas9 can be useful for genome editing, its main limitation is its relatively long PAM, which restricts the number of editable sites in any given genomic locus.

In some embodiments, the present invention contemplates shorter and less stringent PAMs for type II-C Cas9 orthologs including, but not limited to, Nme2Cas9. Although it is not necessary to understand the mechanism of an invention, it is believed that short and less stringent PAMs partially relieve target restriction limitations, while still leaving many, if not most, of the advantages of Nme1Cas9 including, but not limited to, small size (e.g., compactness) for efficient all-in-one AAV delivery and improved target accuracy (e.g., reductions in off-target cleavages). In addition, minimized sgRNAs for Nme1Cas9 discussed above are also compatible with Nme2Cas9 constructs. Consequently, such truncated guide RNAs could likely be used for genome editing with Nme2Cas9 as well.

In one embodiment, the present invention contemplates an HpaCas9 PAM having a sequence of NNNNGNTTT. Despite the fact that the long PAM limits the number of targetable sites in the human genome it is believed that the HpaCas9 PAM may target sites with very high accuracy that is similar to the extreme accuracy Nme1Cas9 (*supra*).

The data presented herein demonstrates the ability of type II-C Cas9 nucleases targeted to short C-rich PAMs to perform genome editing in human (HEK293T) cells. Certo et al., "Tracking genome engineering outcome at individual DNA breakpoints" *Nature Methods* 8:671-676 (2011). For example, HpaCas9 and Nme2Cas9, were shown to provide efficient genome

editing at specific loci demonstrating that they are active in mammalian cells. Figure 19 and Table 2.

Table 2: Representative Type II-C Cas9 Orthologs Target Sequences in The Human Genome

Cas9	Spacer sequence	PAM	Chromosome
Nme2	GAATATCAGGAGACTAGGAAGGAG	GAGGCCTA	19
Hpa	GGACAGGAGTCGCCAGAGGCCGGT	GGTGGATTT	4
Smu	GCACCTGCCTCGTGGAATACGGT	AAACCTAC	Traffic Light Reporter

These data show that both Nme2Cas9 and HpaCas9 performed genome editing at comparable levels to the previously validated Nme1Cas9 at the same genomic locus. For SmuCas9, the efficiency of editing is relatively low, though it is significant that the activity is not zero, and efficiency improvements are expected. Nme2Cas9 was then used to test fourteen (14) additional sites in the traffic light reporter (TLR) integrated into the genome of HEK293T cells. In these assays, each site conforms to a PAM template that a "C" is the fifth nucleotide of the PAM region (i.e., NNNNCNNN). Remarkably, all fourteen sites were edited by Nme2Cas9, indicating that this enzyme is consistently active with a variety of guides in mammalian cells. The most successful guide RNAs conform to the NNNNCCN PAM consensus. Figure 20.

Type II-C Cas9 ortholog cleavage was tested for sensitivity to anti-CRISPR proteins. Anti-CRISPR proteins are naturally occurring proteins that can turn Cas9 off when Cas9 activity is no longer desired. The data show that all three Type II-C Cas9 orthologs are inhibited by certain anti-CRISPRs. Figure 21. The controllability of these Cas9 orthologs by anti-CRISPRs could increase their potential utility in genome editing.

B. Nme2Cas9 Gene Editing

The data presented herein shows gene editing using the Nme2Cas9-sgRNA complex. The data employs the traffic light reporter (TLR) system to demonstrate that any CC dinucleotide in a gene target sequence can function as a PAM, within the context of an NNNNCC sequence (supra). Figure 22. Blue bars are the % of cells that exhibit fluorescence, whereas red bars indicate % editing more accurately based on sequencing ("TIDE analysis"). These data confirm that a dinucleotide is sufficient for Nme2Cas9 PAM binding as opposed to a requirement for a trinucleotide sequence (e.g, the "X" in the sequence NNNNCCX). Although it is not necessary

to understand the mechanism of an invention, it is believed that this means that Nme2Cas9 editable genomic target sites are at least as frequent as SpyCas9 editable sites, and more frequent than with SauCas9, Nme1Cas9 or CjeCas9 and other current alternatives.

Furthermore, T7E1 assays were employed to analyze editing of native genomic sites (e.g., not an integrated, artificial fluorescent reporter). These data suggest that, in some situations, the second "C" might not even be required. See, Figure 23. Note that target sites DeTS1 and DeTS4, both in the AAVS1 locus, enables editing at target sites with NNNNCA and NNNNCG candidate PAMs, respectively. Several of these Nme2Cas9 target sites are disclosed herein. See, Table 3.

Table 3: Representative PAM Target Sites For Nme2Cas9

Target site name	Target locus	Target Sequence (N's-PAM)
Nme2TS1	AAVS1	ATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGG
Nme2TS4	AAVS1	CAGATAAGGAATCTGCCTAACAGGAGGTGGGGGTTAGACGAATATCAGGAGA
Nme2TS5	AAVS1	GGGGTTAGACGAATATCAGGAGACTAGGAAGGAGGAGGCCTAAGGATGGGGG
Nme2TS6	Chr. 14	CCCACCCGGCGGCGCCTCCCTGCAGGGCTGCTCCCCAGCCCAAACCGCCGCG
Nme2TS10	AAVS1	TCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGCCCTATGTCCACTTC
Nme2TS11	AAVS1	TGGGTACTTTTATCTGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGG
Nme2TS12	AAVS1	GTAGGGGAGCTGCCCAAATGAAAGGAGTGAGAGGTGACCCGAATCCACAGGA
Nme2TS13	AAVS1	TAGCACCTCTCCATCCTCTTGCTTTCTTTGCCTGGACACCCCGTTCTCCTGT
Nme2TS14	AAVS1	GTCTCCCTTGCGTCCCGCCTCCCCTTCTTGTAGGCCTGCATCATCACCGTTT
Nme2TS15	AAVS1	CCTCACCCAACCCCATGCCGTGTTCACTCGCTGGGTTCCCTTTTCCTTCTCCT
Nme2TS16	Chr. 14	GCGCAGGACAGGAGTCGCCAGAGGCCGGTGGTGGATTTCCTCCCCGCATCTC
Nme2TS17	Chr. 14	CGCGGGGACGCCCAGCGGCCGGATATCAGCTGCCACGCCCGCGTGGGCGGA
Nme2TS22	VEGF	GATTCCAATAGATCTGTGTGTCCCTCTCCCCACCCGTCCCTGTCCGGCTCTC
Nme2TS23	VEGF	TGACCCCTGGCCTTCTCCCCGCTCCAACGCCCTCAACCCCA

		CACGCACACAC
Nme2TS24	VEGF	TCCCTCTCCCCACCCGTCCCTGTCCGGCTCTCCGCCTTCCCC TGCCCCCTTC
Nme2TS25	VEGF	ACACGCACACACTCACTCACCCACACAGACACACACGTCC TCACTCTCGAAG
Nme2TS26	Chr. 7 (CFTR)	TAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCT GGATTATGCCT
Nme2TS27	Chr. 7 (CFTR)	TTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCAT TAAAGAAAAT

Although it is not necessary to understand the mechanism of an invention, it is believed that these data suggest that there may be candidate editing sites in a genome at every 4-8 base pairs, on average. These data also suggest that most Cas9 sgRNAs have some functionality, consequently the need for sgRNA screening may be overemphasized in the art.

C. Rapidly-Evolving PAM-Interacting Domains

In vivo applications of CRISPR-Cas9 have the potential to transform many areas of biotechnology and therapeutics. There are thousands of Cas9 orthologs in nature, only a handful of which have been validated for *in vivo* genome editing. The Cas9 from *Streptococcus pyogenes* (SpyCas9) has been widely used due to its high efficiency and non-restrictive NGG protospacer adjacent motif (PAM). However, the relatively large size of SpyCas9 restricts its use in *in vivo* therapeutic applications using delivery shuttles with limited packaging capacity such as adeno-associated virus (AAV). Several smaller Cas9 orthologs are known to be active in mammalian cells, but they possess more restrictive PAMs that limit target site density. The natural variation in the PAM Interacting Domains (PIDs) of closely related Cas9 orthologs may be taken advantage of to identify a genome editing enzyme that overcomes these limitations. In some embodiments, the present invention contemplates using an Nme2Cas9 complex which is compact, naturally hyper-accurate Cas9 with an N_4CC PAM. The data presented herein show that Nme2Cas9 is a high-fidelity mammalian genome editing platform that affords the same target site density as SpyCas9. Delivery of Nme2Cas9 with its guide RNA via an all-in-one AAV vector leads to efficient genome editing in adult mice, with *Pcsk9* gene targeting in the liver inducing serum cholesterol reduction with no significant off-targeting (*infra*). Nme2Cas9 also provides a unique combination of all-in-one AAV compatibility, natural hyper-accuracy, and high target site density for in vivo genome editing in mammals.

In addition to target density, minimizing off-target activity (e.g., cleavage at undesired loci) of a Cas9 is highly desirable for its use as a safe therapeutic agent. Wild-type (wt) SpyCas9 possesses a high degree of off-target activity due to its unique hybridization kinetics. (Klein et al, 2018). In particular, questions remain regarding their on-target editing efficiency and these variants do not overcome the above discussed limitations regarding overall size. In contrast, it has been shown herein that embodiments of Nme1Cas9 and CjeCas9 comprise naturally accurate gene editing activity. Although it is not necessary to understand the mechanism of an invention, it is believed that no Cas9 ortholog has been previously reported that: i) is active in human cells; ii) exhibits the exceptionally high target-site density of SpyCas9; iii) is sufficiently compact for all-in-one AAV deliverability; and iv) is naturally hyper-accurate. In one embodiment, the present invention contemplates an Nme2Cas9 as a genome editing platform comprising all of the characteristics described above. For example, Nme2Cas9 comprises a binding site comprising a high affinity for an N_4CC PAM, is hyper-accurate and functions efficiently in mammalian cells. In one embodiment, Nme2Cas9 is packaged in an all-in-one AAV delivery platform for therapeutic genome editing.

1. Closely-Related Nme1Cas9 Orthologs With Rapidly-Evolving PIDs

It has previously been reported that Nme1Cas9 (from *Neisseria meningitidis* strain 8013) is a small, hyper-accurate Cas9 for *in vivo* genome editing (Amrani et al, 2018). However, Nme1Cas9 binds to a long PAM (N_4GMTT) which limits its use in certain contexts where a small window can be targeted. PAM recognition by Cas9 occurs predominantly through protein-DNA interaction between the PAM-Interacting Domain (PID) of Cas9 and the nucleotides adjacent to the PAM. PIDs are subject to high selection pressure by phages and other mobile genetic elements (MGEs). For example, anti-CRISPR proteins have been shown to interact with PIDs to inhibit Cas9 (*infra*). This may result in closely-related Cas9 orthologs having PIDs that recognize drastically different PAMs.

Recently, this principle was highlighted using two species of *Geobacillus*. *G. sterothermpophilus's* was determined to comprise a PID specific for a N_4CRAA PAM but when exchanged for a strain *LC300* PID its affinity changed to a N_4GMAA PAM (Harrington et al, 2017). It was hypothesized that given that *N. meninigitidis* strains are highly sequenced, a closely related Cas9 ortholog could be found with rapidly-evolved PIDs that recognize different PAMs. Cas9 orthologs with high sequence identity (>80%) to NmeCas9 strain 8013 were

investigated because this Cas9 has been fully characterized for genome editing, is small and hyper-accurate. Several Cas9 orthologs were identified which differed in their PID amino acid sequences a compared with strain 8013. Figure 34A.

Three distinct groups of Cas9 orthologs were found with drastically different PIDs. Figure 35A. One strain was selected from each PID group, for example, De11444 from group 2 and 98002 from group 3. These two CRISPR loci had intact Cas9 open reading frames and CRISPR arrays with several spacers, which suggest they are active loci. Interestingly, the crRNA and tracrRNA of these CRISPR loci were identical to that of 8013 and can utilize the same sgRNAs. Figure 35B.

To test whether these Cas9 orthologs indeed had PIDs with affinity for different PAMs, because of the high sequence identity in the remainder of the protein from these orthologs, the 8013 PID was interchanged with the 98002 PID and the De11444 PID. To identify the PAMs, these protein "chimeras" were recombinantly expressed, purified and used for *in vitro* PAM identification as described previously. Briefly, a DNA fragment comprising a protospacer and a ten (10) nucleotide randomized sequence downstream was cleaved *in vitro* using recombinant Cas9 and an sgRNA targeting the protospacer. Figure 34B. A G23 nucleotide spacer length was used for the sgRNA, consistent with Nme1Cas9 8013 and other type II-C systems studied. The PAM identification assay revealed that these different Cas9 chimeras had PIDs recognizing different PAMs. For example, by recognizing a C residue at position 5 instead of a G recognized by Nme1Cas9 8013 with its N_4GATT PAM. Figure 34C.

However, the remaining nucleotides could not be confidently characterized due to the low cleavage efficiency of the chimeric proteins, which suggests that the few residues outside of the PID are likely involved for efficient activity. Figure 35C. To further resolve the PAMs, an *in vitro* assay was performed on a library with a 7-nucleotide randomized PAM, with a C at position 5 (e.g., NNNNCNNN). The results suggested that NmeCas9-De11444 and NmeCas9-98002 recognized NNNNCC(A) and NNNNCAAA PAMs, respectively. Figure 35D. NmeCas9-De11444 had a strong preference for the C at position 5, but less so for nucleotides 6 and 7. As used herein, the Cas9 De11444 ortholog is termed "Nme2Cas9", and the Cas9 98002 ortholog is termed "Nme3Cas9".

We also performed this assay using full-length (e.g., not PID-swapped) Nme2Cas9 and observed similar results. Figure 34E. These results suggest that Nme2Cas9 and Nme3Cas9 have PIDs recognizing drastically different PAMs than that of Nme1Cas9.

2. Nme2Cas9 In Human Cells

Because the Nme2Cas9 PID binds with a small PAM sequence, this ortholog is useful for human genome editing, especially when high-targeting density is involved. To characterize the Nme2Cas9, a full-length (not PID-swapped) humanized Nme2Cas9 was cloned into a CMV-driven plasmid along with NLSs for mammalian expression. For characterization in human cells, a Traffic Light Reporter system was used similar to the one described previously (Certo *et al.*, 2011)

Induction of +1 frameshift indels were created by imperfect repair via non-homologous end joining (NHEJ) in the TLR 2.0 locus. In the absence of a donor DNA an in-frame mCherry protein resulted, which can be quantified through flow cytometry. Figure 36A. As an initial test, a Nme2Cas9 plasmid was transfected along with fifteen (15) sgRNA plasmids with spacers targeting protospacers with N_4CCX PAMs. As controls, SpyCas9 and Nme1Cas9 were used along with their cognate sgRNAs targeting NGG and N_4GATT protospacers, respectively. Cells were harvested after seventy-two (72) hours and the number of mCherry positive cells was quantified for each target site. SpyCas9 and Nme1Cas9 showed efficient editing at their respective targets (~28% and 10% mCherry, respectively) Figure 36B. For Nme2Cas9, all fifteen (15) targets with N_4CCX PAMs were functional to various degrees (ranging from 4% to 20% mCherry), while NmeCas9 treatments without accompanying sgRNA and/or N_4GATT controls yielded no mCherry cells. Figure 36B. These data suggested that Nme2Cas9 recognizes an N_4CC PAM in human cells.

To further resolve Nme2Cas9 PAMs, target sites were also tested with N_5CX and N_4CD (D = A, T, G) in TLR reporter cells. No detectable editing was observed at target sites with N_5CX and N_4CD PAMs, suggesting that both C nucleotides at positions 5 and 6 are required for Nme2Cas9's activity based on the TLR 2.0 reporter. Figures 37A and 37B. These results demonstrate that Nme2Cas9 comprises a PID that binds to an N_4CC PAM and is consistently functional in mammalian cells at the TLR 2.0 locus.

The length of the spacer portion of the crRNA differs between different Cas9 orthologs. SpyCas9's optimal spacer length is twenty (20) nucleotides, however, truncations down to

seventeen (17) nucleotides are tolerated. Fu et al., Nature Biotechnology 32, 279 (2014). In contrast, Nme1Cas9 comprises sgRNAs with twenty-four (24) nucleotide spacers and tolerates truncations down to eighteen (18) nucleotides. (Amrani et al., 2018). To test the spacer length for Nme2Cas9, sgRNA plasmids were created that targeted the same locus, but with varying spacer lengths. Figure 36C and Figure 37B. Comparable activities were observed when G23, G22 and G21 spacers were used, with a significant decrease in activity when the guide was truncated to G20 and G19. Figure 36C. These results suggest that Nme2Cas9's optimal spacer length is between 22-24 nucleotides, similar to that of Nme1Cas9, GeoCas9 and CjeCas9. Therefore, all experiments described below were performed with 23-24 nucleotide spacers.

Cas9 orthologs are believed to use their HNH and RuvC domains to induce a double stranded break in the complementary and non-complementary strands of the target DNA, respectively. Alternatively, Cas9 nickases have been used to improve genome editing specificity and homology-directed repair (HDR) by creating overhangs. (Ran et al, 2013). However, this approach has only been successful by use of SpyCas9 due to its high target density. To use Nme2Cas9 as a nickase, Nme2Cas9^{D16A} and Nme2Cas9^{H588A} were created which provide mutations in the catalytic residues of the RuvC and HNH domains, respectively. Since TLR 2.0 can also be used to study the efficiency of HDR, where a repaired locus expresses GFP when a donor is provided, a donor DNA sequence was included to test HDR with these Nme2Cas9 nickases. Target sites were selected within the TLR 2.0 gene to test the functionality of each nickase using guide RNAs that targeted cleavage sites spaced 32 bp and 64 bp apart. As a control, wild type Nme2Cas9 targeted to a single site showed efficient editing, accompanied by induction of both NHEJ and HDR repair pathways. For nickases, the cleavage sites spaced 32 bp and 64 bp apart showed editing using the Nme2Cas9^{D16A} (HNH nickase), but neither target was nicked using Nme2Cas9^{H588A}. Figure 36D.

Cas9 orthologs comprise a seed sequence that usually hybridizes to a target sequence between eight to twelve (8-12) nucleotides proximal to the PAM. Mismatches (e.g., non-complementarity) between the seed sequence and the PAM can reduce Cas9 nuclease activity. A series of transient transfections were performed that targeted the same locus in the TLR 2.0 gene by walking single nucleotide mismatches along a twenty-three (23) nucleotide spacer. Figure 37C. Similar to other Cas9 orthologs, the data suggest that Nme2Cas9 possesses a "seed sequence" in the first eight-to-nine (8-9) nucleotides that hybridize to a target sequence proximal

to the PAM, as deduced from the decrease in the number of mCherry positive cells. Even though tolerance to mismatches is highly dependent on the sequence and the target locus of an sgRNA, these results suggest that Nme2Cas9 has very low tolerance for mismatches particularly in its seed sequence.

3. Nme2Cas9 Genome Editing Efficiency

Nme2Cas9 was used to target forty (40) different target sites throughout the human genome in HEK293T cells using transient transfections. Table 4.

Table 4: Representative HEK293T Cell Nme2Cas9 Target Sites

Number	Site Name	Genomic Seq	PAM	Locus	% Mod. Freq.	TIDE Primer Name	Fw TIDE Primer	Rv TIDE Primer
1	TS1	GGTTCTGGGTACTTTTATCTGTCC	CCTCCAGC	AAVS1	0.2	AAVS1_TIDE1	TGGCTTAGCAGCTCTTCCAT	AGGAGTCAGGAGCAACTTATTCTG
2	TS4	GTCTTGCCTAACAGGAGGTGGGGGT	TAGACCAA	AAVS1	11	AAVS1_TIDE1	TGGCTTAGCAGCTCTTCCAT	AGGAGTCAGGAGCAACTTATTCTG
3	TS5	GAATATCAGGAGACTAGGAAGGAG	GAGGCATA	AAVS1	15	AAVS1_TIDE1	TGGCTTAGCAGCTCTTCCAT	AGGAGTCAGGAGCAACTTATTCTG
4	TS6	GCCTCCCTGCAGGGCTGCTCCC	CAGCCCAA	LINC01588	20	LINC01568_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
5	TS15	GAGCTAGTCTTCTTCCTCCAAGCC	GGGCCATA	AAVS1	3.5	AAVS1_TIDE1	TGGCTTAGCAGCTCTTCCAT	AGGAGTCAGGAGCAACTTATTCTG
6	TS11	GATCTGTCCCCTCCACCCCACAGT	GGGGCCAC	AAVS1	9	AAVS1_TIDE1	TGGCTTAGCAGCTCTTCCAT	AGGAGTCAGGAGCAACTTATTCTG
7	TS12	GGCCCAAATGAAGGAGTGAGAGG	TGACCCAA	AAVS1	10	AAVS1_TIDE1	TGGCTTAGCAGCTCTTCCAT	TAGGAAGGAGGAGGAGGCCTAAG
8	TS13	GCATCCTGTTGCTTTGTTTGCCTG	GACACAGG	AAVS1	2	AAVS1_TIDE1	TGCATCCTCCTCCTCACTCC	TAGGAAGGAGGAGGAGGCCTAAG
9	TS16	GGAGTCGCCAGAGGCCGGTGGTGG	ATTTCCTC	LINC01568	26	LINC01588_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
10	TS17	GCCCAGCGGCGGGGATATGAGCTGC	CACGCCCG	LINC01588	0.2	LINC01568_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
11	TS18	GGAAGGGAACATATTACTATTGC	TTTCCCTC	CYBB	1	NT555_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
12	TS19	GTGGAGTGGCCTGCTATCAGCTAC	CTATCCAA	CYBB	5	NT555_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
13	TS20	GAGGAAGGAACATATTACTATTG	CTTTCCAT	CYBB	11.2	NT555_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
14	TS21	GTGAATTCTCATCAGCTAAAATGC	CAAGCCTT	CYBB	1	NT555_TIDE	TAGAGAACTGGGTAGTGTG	CCAATATTGCATGGGATGG
15	TS25	GCTCACTCACCCACACAGACACAC	ACCTCCTC	VEGF A	15.6	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
16	TS26	GGAAGAATTTCATTCTGTTCTCAG	TTTTCCTG	CFTR	2	hCFTR_TIDE1	TGGTGATTATGGGAGAACTGGAGC	ACCATTGAGGACGTTTGTCTTCAC
17	TS27	GCTCAGTTTTCCTGGATTATGCCT	GGCACCAT	CFTR	4	hCFTR_TIDE1	TGGTGATTATGGGAGAACTGGAGC	ACCATTGAGGACGTTTGTCTTCAC
18	TS31	GCGTTGGAGCGGGGAGAAGGCCAG	GGGTCAAT	VEGFA	9	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
19	TS34	GGGCCGCGGAGATAGCTGCAGGGC	GGGTCCCC	LINC01588	3	LINC01568_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
20	TS35	GCCCACCCGGCGGCGCCTCCCTGC	AGGGCTGC	LINC01588	5	LINC01568_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
21	TS36	GCGTGGCAGCTGATATCCGGCGC	TGGGCCTC	LINC01588	5	LINC01568_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
22	TS37	GCCGCGCCGCGACGTGGAGCCAGC	CCCGCAAA	LINC01568	0.5	LINC01588_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
23	TS38	GTGCTCCCAGCCCAAACCGCCGC	GGCGCAAG	LINC01568	2	LINC01588_TIDE	AGAAGGAGCCTTCTGACTGCTGCAGA	ATGACAGACACAACCAGAGGGCA
24	TS41	GTCAGATTGGCTTGCTCGGAATTG	CCAGCCAA	AGA	3	AGA_TIDE1	GGCATAAGGGAATCCAAGGTC	CATGTCCTCAAGTCAAGAACAAG
25	TS49	GCTGGGTGAATGGAGCGAGCAGCG	TCTTCGAG	VEGFA	3	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
26	TS45	GTCCTGGAGTGACCCCTGGCCTTC	TGCCCACT	VEGFA	7.4	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
27	TS46	GATCCTGGAGTGACCCCTGGCCTT	CTCCCACC	VEGFA	9	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
28	TS47	GTGTGTCCCTCTCCCCAGCCCGTCC	CTGTCCCC	VEGFA	23.1	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
29	TS48	GTTGGAGCGGGGAGAAGGCCAGGG	GTCACTCC	VEGFA	2	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
30	TS42	GCGTTGGAGCGGGGAGAAGGCCAG	GGGTCAAT	VEGFA	4	VEGF_TIDE3	GTACATGAAGCAACTCCAGTCCCA	ATCAAATTCCAGCACCGAGCGC
31	TS50	GTACCCTGCAATAATTTGGCTGGC	AATTCCAA	AGA	8	AGA_TIDE1	GGCATAAGGGAATCCAAGGTC	CATGTCCTCAAGTCAAGAACAAG
32	TS51	GATAATTTGGCTGGCAATTCCGAG	CAAGCCAA	AGA	4.5	AGA_TIDE1	GGCATAAGGGAATCCAAGGTC	CATGTCCTCAAGTCAAGAACAAG
33	TS58 (DS11)	GCAGGGCCCAGCTGTGCTTCTCTG	GGGCCCTC	VEGFA	5	VEGF_TIDE5	ACACGGGCAGCATGGGGATAGTC	GCTGAGGGAGGAGTCCCACTGTCCA
34	TS59 (DS12)	GGATGGCAGGCGGAAGTTGTACTG	GGGCCCAG	VEGFA	11.5	VEGF_TIDE5	CCTGTGTGGCTTTGCTTTGGTGG	GTAGGGTGTGATGGGAGGCTAAGC
35	TS60 (DS12)	GGGTGAGAGAGTGAGAGAGAGAA	CGGGCCAG	VEGFA	3	VEGF_TIDE5	CCTGTGTGGCTTTGCTTTGGTGG	GTAGGGTGTGATGGGAGGCTAAGC
36	TS61 (DS14)	GTGAGCAGGCACGTGTGCCAACAT	GGGCCCAG	VEGFA	9.5	VEGF_TIDE5	CCTGTGTGGCTTTGCTTTGGTGG	GTAGGGTGTGATGGGAGGCTAAGC
37	TS62 (DS15)	GCGTGGGGGGCTCCGTGCCCAACGC	GGGTCCAT	VEGFA	3.4	VEGF_TIDE5	GGAGGGAGAGTAGCTCCGCCGAGG	AGACCGGGTGGCAGTGACAGCAAG
38	TS63 (DS16)	GCATGGGCAGGGGCTGGGGTGCAC	AGGCCCAG	VEGFA	16	VEGF_TIDE5	AGGGAGAGGGGAAGTGTGGGGAAGG	CTCTTCCTGCTGTGTGGGCACGAC
39	TS64	GAAAATTGTGAATTCCAGATCCAC	AAGCCCAA	FANCJ	7	Fanc_TIDE5	GTTGGGGGCTCTAAGTTATGTAT	CTTCATCTGTATCTTCAGGATCA
40	TS65	GAGCAGAAAAAATTGTGAATTCC	AGATCCAC	FANCJ	5	Fanc_TIDE5	GTTGGGGGCTCTAAGTTATGTAT	CTTCATCTGTATCTTCAGGATCA

72-hours post transfection, cells were harvested followed by gDNA extraction and selective amplification of the targeted locus. A Tracking of Indels by Decomposition (TIDE) analysis was used to measure indel rates at each locus. Efficient editing by Nme2Cas9 was observed, even though indel rates varied significantly depending on the target sequence and the locus. Figure 38A. Moreover, Nme2Cas9's affinity for target sites near/at therapeutically-relevant loci such as *CYBB* (mutations cause x-linked chronic granulomatous disease) and *AGA* (mutations cause

aspartylglycosaminuria) suggests Nme2Cas9 has therapeutic potential. In addition, editing efficiency could be increased by increasing the quantity of the Nme2Cas9 plasmid. Figure 39A. Taken together, these results demonstrate that Nme2Cas9 can be constructed to selectively edit specific target genomic sites in HEK293T cells.

In addition to HEK293T cells, Nme2Cas9's gene editing efficiency was determined in several other mammalian cells, including human leukemia K562 cells, human osteosarcoma U2OS cells and mouse liver hepatoma Hepa1-6 cells. A lentiviral construct expressing Nme2Cas9 was created and transduced K562 cells to stably express Nme2Cas9 under the control of SFFV promoter. This stable cell line did not show any significant differences with respect to growth and morphology as compared to untreated cells, suggesting Nme2Cas9 is not toxic when stably expressed. These cells were transiently electroporated with plasmids expressing sgRNAs targeting several target sites and analyzed after seventy-two (72) hours for indel rates by TIDE. Efficient editing was observed at the three sites tested, demonstrating Nme2Cas9's ability to function in K562 cells. For Hepa1-6 cells, plasmids encoding Nme2Cas9 and sgRNA were co-transfected using techniques similar to HEK293T transduction described above. These data also show that Nme2Cas9 efficiently edited *Pcsk9* and *Rosa26* sites in this mouse cell line. Figure 38B.

Previous work suggests that ribonucleoprotein (RNP) delivery of Cas9s, instead of plasmid transfection, may be an alternative choice for some genome editing applications. For example, off-target effects of SpyCas9 may be significantly reduced with RNP electroporations compared to plasmid delivery. Kim et al., *Genome Research* 24:1012-1019 (2014). To test whether Nme2Cas9 is functional by RNP delivery, a His-tagged Nme2Cas9 was cloned along with three (3) nuclear localization signals (NLSs) and a purified recombinant protein into a bacterial expression construct. sgRNAs targeting several validated target sites were generated by T7 *in vitro* transcription. Electroporation of a Nme2Cas9:sgRNA complex induced successful editing at the target sites, as detected by TIDE. Figure 38C. These results suggest that Nme2Cas9 can be delivered as a plasmid, or as an RNP complex. Overall, these results demonstrate that Nme2Cas9 is functional in various cell types with different modes of delivery.

4. Anti-CRISPR Protein Inhibition

Five (5) anti-CRISPR (Acr) protein families against Nme1Cas9 from diverse bacterial species have been reported to inhibit Nme1Cas9 *in vitro* and in human cells. (Pawluk et al. 2016,

Lee et al., *mBio*, in press). Considering the high sequence identity between Nme1Cas9 and Nme2Cas9, it seemed likely that at least some species within these Acr families might also inhibit Nme2Cas9. All five Acr families were recombinantly expressed, purified and Nme2Cas9's ability to cleave a target sequence *in vitro* was tested (10:1 Acr:Cas9 molar ratio). As a negative control, an inhibitor for the type I-E CRISPR system in *E. coli* (AcrE2) was used. As expected, all Arc families inhibited Nme1Cas9, while AcrE2 failed to do so. In particular, Acrs IIC1$_{Nme}$, -IIC2$_{Nme}$, -IIC3$_{Nme}$ and -IIC4$_{Hpa}$ inhibited Nme2Cas9 gene editing activity. Figure 40A, top.

Strikingly, AcrIIC5$_{Smu}$ did not inhibit Nme2Cas9 *in vitro* even at 10-fold excess, suggesting that it likely inhibits Nme1Cas9 by interacting with a PID. To further confirm this, the same *in vitro* cleavage assay was performed using a hybrid version of NmeCas9 (e.g., Nme1Cas9 with the PID of Nme2Cas9). Due to the reduced activity of this hybrid, higher concentration (~30X) of Cas9 was used to achieve similar cleavage profile while maintaining the 10:1 Cas9:Acr molar ratio. Consistent with the initial results, no inhibition by AcrIIC5$_{Smu}$ on this protein chimera was observed. Figure 41.. The inability of AcrIIC5$_{Smu}$ to inhibit the hybrid protein further suggests that AcrIIC5$_{Smu}$ likely interacts with the PID of Nme1Cas9.

The above *in vitro* data, suggested that Acrs -IIC1$_{Nme}$, -IIC2$_{Nme}$, -IIC3$_{Nme}$ and -IIC4$_{Hpa}$ could be used as off-switches for Nme2Cas9 genome editing. To test this, transfections were performed as described above in the presence or absence of plasmids encoding Acrs driven by mammalian promoters. Approximately 150 ng of each plasmid (e.g., having a 1:1:1 ratio of sgRNA:Cas9:Acr) was transfected, as most ACRs have been reported to inhibit Nme1Cas9 at those ratios. (Pawluk et al., 2016). As expected from the *in vitro* experiment, AcrIIC1$_{Nme}$, -IIC2$_{Nme}$, -IIC3$_{Nme}$ and -IIC4$_{Hpa}$ inhibited Nme2Cas9 genome editing, while AcrIIC5$_{Smu}$ failed to do so. (Figure 40B. Moreover, complete inhibition was observed to be below detection levels by Acr3Nme and Acr4Hpa, suggesting their high potency as compared to AcrsIIC1$_{Nme}$ and AcrIIC2$_{Nme}$. To further compare the potency of AcrIIC1$_{Nme}$ and AcrIIC4$_{Hpa}$, experiments were performed at various ratios of Acr to Cas9. Figure 40C. Consequently, AcrIIC4$_{Hpa}$ is a highly potent inhibitor against Nme2Cas9, with concentrations as low as 25ng:100ng Acr:Cas9 inhibiting Nme2Cas9 by 4 fold. Together, these data suggest that Acr proteins can be used as off-switches for Nme2Cas9-based applications.

5. Nme2Cas9 Hyper-Accuracy

Off-target effects could potentially confound therapeutic applications during *ex vivo* and *in vivo* human gene therapy by creating unintended mutations. Since wildtype SpyCas9 has a relatively high number of off-target sites in human cells, there have been several efforts to engineer high-fidelity SpyCas9 variants with variable success. In contrast, Nme1Cas9 is naturally hyper-accurate, demonstrating remarkable fidelity in cells and mouse models. Previous work shows that hybridization kinetics, which is not determined by the PID, may determine the fidelity of a Cas9, therefore suggesting that Nme2Cas9 may also be hyper-accurate.

To empirically assess NmeCas9 off-target profiles, Genome-Wide, Unbiased Identification of double-stranded breaks Enabled by Sequencing (GUIDE-Seq) techniques were used to determine potential off-target sites in an unbiased fashion. GUIDE-Seq relies on the incorporation of double-stranded oligodeoxynucleotides (dsODNs) into DNA double-stranded break sites throughout the genome. These cleavage sites are detected by amplification and high-throughput sequencing.

As a benchmark for GUIDE-Seq, wildtype SpyCas9 was used. In particular, SpyCas9 and Nme2Cas9 were able to be cloned into identical backbones driven by the same promoter, and used to target the same sites because of their non-overlapping PAMs. This technique allows side-by-side comparison the two nucleases. Six (6) dual sites (DS) were targeted in *VEGFA* with a NGGNCCN sequence. Figure 42A. Seventy-two (72) hours after transfection, TIDE analysis was performed on the target sites. Nme2Cas9 induced indels at all six (6) sites, albeit at low efficiencies at two of them, while SpyCas9 induced indels at 4/6 sites. Figure 42B. On two of those 4 sites (DS1 and DS4), SpyCas9 induced ~7 fold more indels than Nme2Cas9, while Nme2Cas9 induced by ~3 folds increase in indels at DS6. For GUIDE-seq, targets DS2, DS4 and DS6 were selected to determine off-target cleavage at sites where Nme2Cas9 is as efficient, less efficient or more efficient than SpyCas9, respectively.

In addition to the three dual target sites, a TS6 target site with a 30-50% indel rate (depending on the cell type) along with the mouse *Pcsk9* and *Rosa26* genes were subjected to GUIDE-Seq analysis. It was considered that the off-target profiles would be more prominent because the TS6 target is known to undergo highly efficient gene editing. In addition, testing of the mouse *Pcsk9* and *Rosa26* sites would then reveal the fidelity of Nme2Cas9 in a different cell line, and candidate loci for *in vivo* genome editing. Consequently, transfections

were performed for each Cas9 along with their cognate sgRNAs and the dsODNs and GUIDE-Seq libraries were prepared. GUIDE-Seq analysis demonstrated efficient on-target editing with both Cas9 orthologs with similar patterns observed by TIDE. For off-target identification, the analysis revealed that while the three SpyCas9 sites had the expected high number of off-target sites (e.g., ranging between approximately between 10 - 1000). Nme2Cas9 had a strikingly clean off-target profile. Specifically, Nme2Cas9 targeting the same dual site showed, at most, one off-target site. See, Figure 42C.

To validate the off-target sites detected by GUIDE-seq, 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 (in some instances, 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. With the Rosa26 sgRNA, Nme2Cas9 induced ~1% editing at the Rosa26-OT1 site in Hepa1-6 cells, compared to ~30% on-target editing. Figure 42D.

Next, to enable the use of SpyCas9 as a benchmark for GUIDE-seq, due to the fact that SpyCas9 and Nme2Cas9 have non-overlapping PAMs they 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 binding properties. This enables side-by-side comparisons of off-targeting with sgRNAs that bind the exact same on-target site. Using matched plasmids expressing each Cas9 and their respective sgRNAs, twenty-eight (28) DSs were targeted at multiple loci throughout the human genome. Seventy-two (72) hours after plasmid delivery, a TIDE analysis was performed on the sites targeted by each nuclease. Nme2Cas9 induced indels at nineteen (19) target sites, albeit at low efficiencies (<5%) at four of them, while SpyCas9 induced indels at twenty-three (23) of the target sites, in one case with <5% efficiency. Three dual target sites were recalcitrant to editing by both nucleases. While SpyCas9 is clearly more efficient overall, both enzymes have similar efficiencies at many of the sites, and at two of the seventeen sites that were edited by both nucleases, Nme2Cas9 was more efficient under these conditions. See, Figure 42E.

It is noteworthy that this off-target site has a consensus Nme2Cas9 PAM (ACTCCCT) with only 3 mismatches at the PAM-distal end of the guide-complementary region (i.e. outside of

the seed). See, Figure 42F. These data support and reinforce our GUIDE-seq results indicating a high degree of accuracy for Nme2Cas9 genome editing in mammalian cells.

On- vs. off-target on these sides were compared by targeted amplification of each locus followed by TIDE analysis. Figure 43A. Interestingly, no indels could be detected at those off-target sites for either sgRNA by TIDE, while efficient on-target editing was observed. Furthermore, the read counts for these off-targets were negligible as compared to those observed in the case of SpyCas9 suggesting Nme2Cas9 is highly specific. (Figure 43C, left versus right, respectively). To further corroborate these GUIDE-Seq results, CRISPRseek was used to computationally predict potential off-target sites for two of the most active sgRNAs with highly similar sites in the genome. (Zhu et al., 2014). These were performed with N_4CX PAMs and 2-5 mismatches, mostly in the PAM-distal region. Figure 43D. Taken together, these data suggest that Nme2Cas9 is a high-fidelity nuclease in mammalian cells.

6. Clinical Applications

In one embodiment, the present invention contemplates an Nme2Cas9 complex as the first compact, hyper-accurate Cas9 with a small non-restrictive PAM for therapeutic genome editing by AAV delivery. Although small, previously reported hyper-accurate Cas9 orthologs have longer PAMs than those disclosed herein, thereby restricting their therapeutic use due to limited target sites in a given gene (and off-target profile in the case of SauCas9). This disadvantage is exacerbated in loci where only a specific window can be targeted, or a precise block deletion is required.

The all-in-one AAV delivery platform established herein can be used to target any gene in any tissue. Moreover, Nme2Cas9's hyper-accuracy enables precise editing of the target genes, therefore ameliorating safety concerns raised due to off-target activities previously observed. To this end, Nme2Cas9 has the potential to not only complement existing tools, but to become a preferred choice for therapeutic genome editing by viral delivery.

Furthermore, inhibition of Nme2Cas9 by various Acrs suggest a possible evolutionary pressure imposed on Cas9 to rapidly evolve a particular domain. Specifically, the lack of inhibition of Nme2Cas9 by $AcrIIC5_{Smu}$ raises the possibility that its mechanism of inhibition is through a PID. Considering that $AcrIIC5_{Smu}$ is the most potent inhibitor of Nme1Cas9 to date, it is contemplated herein where $AcrIIC5_{Smu}$ can be used to robustly turn off Nme1Cas9 but not

Nme2Cas9. This is of particular interest in cellular contexts where multiplexing would be enhanced by the ability to control a specific ortholog.

Finally, while there are thousands of Cas9 orthologs in the public database, only a handful of which have been characterized. Some embodiments contemplated herein take advantage of the natural variation in closely-related Cas9 orthologs to create two novel Cas9 nucleases, namely Nme2Cas9 and Nme3Cas9, with N4CC and N4CAAA PAMs, respectively. The data presented herein demonstrate that even closely related orthologs can have vastly different properties. For example, these orthologs use the exact same sgRNA as Nme1Cas9, which circumvent the difficulties in the prediction of tracrRNAs and determining the right spacer length for each ortholog. Furthermore, it is likely that shorter and more stable sgRNAs (such as chemical modifications) can be engineered to expand to all three nucleases. These characteristics may ease genome editing efforts and reduce the costs associated with protein and RNA engineering.

It should be apparent to one of skill in the art that the embodiments described herein are not restricted to Cas9s and can be applied to other Cas proteins such as Cas12 and Cas13. It should also be appreciated that Cas9's hyper-variability is not restricted to PIDs. It is considered herein that strains exist which share high degree of homology with a given Cas9 but differ in other domains due to other types of selective pressure. Taken together, Nme2Cas9 is a novel nuclease which improves the current CRISPR platforms for therapeutic genome editing.

V. Nucleotide Delivery Platforms

Aside from the above described AAV nucleotide delivery systems, the present invention contemplates several delivery systems compatible with nucleic acids that provide for roughly uniform distribution and have controllable rates of release. Some embodiments of the present invention contemplate nucleic acid delivery systems encoding Type II-C Cas9-sgRNA complexes as described herein.

A variety of different media are described below that are useful in creating nucleic acid delivery systems. It is not intended that any one medium or carrier is limiting to the present invention. Note that any medium or carrier may be combined with another medium or carrier; for example, in one embodiment a polymer microparticle carrier attached to a compound may be combined with a gel medium.

Carriers or mediums contemplated by this invention comprise a material selected from the group comprising gelatin, collagen, cellulose esters, dextran sulfate, pentosan polysulfate, chitin, saccharides, albumin, fibrin sealants, synthetic polyvinyl pyrrolidone, polyethylene oxide, polypropylene oxide, block polymers of polyethylene oxide and polypropylene oxide, polyethylene glycol, acrylates, acrylamides, methacrylates including, but not limited to, 2-hydroxyethyl methacrylate, poly(ortho esters), cyanoacrylates, gelatin-resorcin-aldehyde type bioadhesives, polyacrylic acid and copolymers and block copolymers thereof.

Microparticles

One embodiment of the present invention contemplates a nucleic acid delivery system comprising a microparticle. Preferably, microparticles comprise liposomes, nanoparticles, microspheres, nanospheres, microcapsules, and nanocapsules. Preferably, some microparticles contemplated by the present invention comprise poly(lactide-co-glycolide), aliphatic polyesters including, but not limited to, poly-glycolic acid and poly-lactic acid, hyaluronic acid, modified polysacchrides, chitosan, cellulose, dextran, polyurethanes, polyacrylic acids, psuedo-poly(amino acids), polyhydroxybutrate-related copolymers, polyanhydrides, polymethylmethacrylate, poly(ethylene oxide), lecithin and phospholipids.

Liposomes

One embodiment of the present invention contemplates liposomes capable of attaching and releasing nucleic acids as described herein. Liposomes are microscopic spherical lipid bilayers surrounding an aqueous core that are made from amphiphilic molecules such as phospholipids. For example, a liposome may trap a nucleic acid between the hydrophobic tails of the phospholipid micelle. Water soluble agents can be entrapped in the core and lipid-soluble agents can be dissolved in the shell-like bilayer. Liposomes have a special characteristic in that they enable water soluble and water insoluble chemicals to be used together in a medium without the use of surfactants or other emulsifiers. Liposomes can form spontaneously by forcefully mixing phosopholipids in aqueous media. Water soluble compounds are dissolved in an aqueous solution capable of hydrating phospholipids. Upon formation of the liposomes, therefore, these compounds are trapped within the aqueous liposomal center. The liposome wall, being a phospholipid membrane, holds fat soluble materials such as oils. Liposomes provide controlled release of incorporated compounds. In addition, liposomes can be coated with water soluble

polymers, such as polyethylene glycol to increase the pharmacokinetic half-life. One embodiment of the present invention contemplates an ultra high-shear technology to refine liposome production, resulting in stable, unilamellar (single layer) liposomes having specifically designed structural characteristics. These unique properties of liposomes, allow the simultaneous storage of normally immiscible compounds and the capability of their controlled release.

In some embodiments, the present invention contemplates cationic and anionic liposomes, as well as liposomes having neutral lipids. Preferably, cationic liposomes comprise negatively-charged materials by mixing the materials and fatty acid liposomal components and allowing them to charge-associate. Clearly, the choice of a cationic or anionic liposome depends upon the desired pH of the final liposome mixture. Examples of cationic liposomes include lipofectin, lipofectamine, and lipofectace.

One embodiment of the present invention contemplates a nucleic acid delivery system comprising liposomes that provides controlled release of at least one nucleic acid. Preferably, liposomes that are capable of controlled release: i) are biodegradable and non-toxic; ii) carry both water and oil soluble compounds; iii) solubilize recalcitrant compounds; iv) prevent compound oxidation; v) promote protein stabilization; vi) control hydration; vii) control compound release by variations in bilayer composition such as, but not limited to, fatty acid chain length, fatty acid lipid composition, relative amounts of saturated and unsaturated fatty acids, and physical configuration; viii) have solvent dependency; iv) have pH-dependency and v) have temperature dependency.

The compositions of liposomes are broadly categorized into two classifications. Conventional liposomes are generally mixtures of stabilized natural lecithin (PC) that may comprise synthetic identical-chain phospholipids that may or may not contain glycolipids. Special liposomes may comprise: i) bipolar fatty acids; ii) the ability to attach antibodies for tissue-targeted therapies; iii) coated with materials such as, but not limited to lipoprotein and carbohydrate; iv) multiple encapsulation and v) emulsion compatibility.

Liposomes may be easily made in the laboratory by methods such as, but not limited to, sonication and vibration. Alternatively, compound-delivery liposomes are commercially available. For example, Collaborative Laboratories, Inc. are known to manufacture custom designed liposomes for specific delivery requirements.

Microspheres, Microparticles And Microcapsules

Microspheres and microcapsules are useful due to their ability to maintain a generally uniform distribution, provide stable controlled compound release and are economical to produce and dispense. Preferably, an associated delivery gel or the compound-impregnated gel is clear or, alternatively, said gel is colored for easy visualization by medical personnel.

Microspheres are obtainable commercially (Prolease®, Alkerme's: Cambridge, Mass.). For example, a freeze dried medium comprising at least one therapeutic agent is homogenized in a suitable solvent and sprayed to manufacture microspheres in the range of 20 to 90 μm. Techniques are then followed that maintain sustained release integrity during phases of purification, encapsulation and storage. Scott et al., Improving Protein Therapeutics With Sustained Release Formulations, Nature Biotechnology, Volume 16:153-157 (1998).

Modification of the microsphere composition by the use of biodegradable polymers can provide an ability to control the rate of nucleic acid release. Miller et al., Degradation Rates of Oral Resorbable Implants {Polylactates and Polyglycolates: Rate Modification and Changes in PLA/PGA Copolymer Ratios, J. Biomed. Mater. Res., Vol. II:711-719 (1977).

Alternatively, a sustained or controlled release microsphere preparation is prepared using an in-water drying method, where an organic solvent solution of a biodegradable polymer metal salt is first prepared. Subsequently, a dissolved or dispersed medium of a nucleic acid is added to the biodegradable polymer metal salt solution. The weight ratio of a nucleic acid to the biodegradable polymer metal salt may for example be about 1:100000 to about 1:1, preferably about 1:20000 to about 1:500 and more preferably about 1:10000 to about 1:500. Next, the organic solvent solution containing the biodegradable polymer metal salt and nucleic acid is poured into an aqueous phase to prepare an oil/water emulsion. The solvent in the oil phase is then evaporated off to provide microspheres. Finally, these microspheres are then recovered, washed and lyophilized. Thereafter, the microspheres may be heated under reduced pressure to remove the residual water and organic solvent.

Other methods useful in producing microspheres that are compatible with a biodegradable polymer metal salt and nucleic acid mixture are: i) phase separation during a gradual addition of a coacervating agent; ii) an in-water drying method or phase separation method, where an antiflocculant is added to prevent particle agglomeration and iii) by a spray-drying method.

In one embodiment, the present invention contemplates a medium comprising a microsphere or microcapsule capable of delivering a controlled release of a nucleic acid for a duration of approximately between 1 day and 6 months. In one embodiment, the microsphere or microparticle may be colored to allow the medical practitioner the ability to see the medium clearly as it is dispensed. In another embodiment, the microsphere or microcapsule may be clear. In another embodiment, the microsphere or microparticle is impregnated with a radio-opaque fluoroscopic dye.

Controlled release microcapsules may be produced by using known encapsulation techniques such as centrifugal extrusion, pan coating and air suspension. Such microspheres and/or microcapsules can be engineered to achieve desired release rates. For example, Oliosphere® (Macromed) is a controlled release microsphere system. These particular microsphere's are available in uniform sizes ranging between 5 - 500 μm and composed of biocompatible and biodegradable polymers. Specific polymer compositions of a microsphere can control the nucleic acid release rate such that custom-designed microspheres are possible, including effective management of the burst effect. ProMaxx® (Epic Therapeutics, Inc.) is a protein-matrix delivery system. The system is aqueous in nature and is adaptable to standard pharmaceutical delivery models. In particular, ProMaxx® are bioerodible protein microspheres that deliver both small and macromolecular drugs, and may be customized regarding both microsphere size and desired release characteristics.

In one embodiment, a microsphere or microparticle comprises a pH sensitive encapsulation material that is stable at a pH less than the pH of the internal mesentery. The typical range in the internal mesentery is pH 7.6 to pH 7.2. Consequently, the microcapsules should be maintained at a pH of less than 7. However, if pH variability is expected, the pH sensitive material can be selected based on the different pH criteria needed for the dissolution of the microcapsules. The encapsulated nucleic acid, therefore, will be selected for the pH environment in which dissolution is desired and stored in a pH preselected to maintain stability. Examples of pH sensitive material useful as encapsulants are Eudragit® L-100 or S-100 (Rohm GMBH), hydroxypropyl methylcellulose phthalate, hydroxypropyl methylcellulose acetate succinate, polyvinyl acetate phthalate, cellulose acetate phthalate, and cellulose acetate trimellitate. In one embodiment, lipids comprise the inner coating of the microcapsules. In these compositions, these lipids may be, but are not limited to, partial esters of fatty acids and hexitiol

anhydrides, and edible fats such as triglycerides. Lew C. W., Controlled-Release pH Sensitive Capsule And Adhesive System And Method. United States Patent No. 5,364,634 (herein incorporated by reference).

In one embodiment, the present invention contemplates a microparticle comprising a gelatin, or other polymeric cation having a similar charge density to gelatin (i.e., poly-L-lysine) and is used as a complex to form a primary microparticle. A primary microparticle is produced as a mixture of the following composition: i) Gelatin (60 bloom, type A from porcine skin), ii) chondroitin 4-sulfate (0.005% - 0.1%), iii) glutaraldehyde (25%, grade 1), and iv) 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC hydrochloride), and ultra-pure sucrose (Sigma Chemical Co., St. Louis, Mo.). The source of gelatin is not thought to be critical; it can be from bovine, porcine, human, or other animal source. Typically, the polymeric cation is between 19,000-30,000 daltons. Chondroitin sulfate is then added to the complex with sodium sulfate, or ethanol as a coacervation agent.

Following the formation of a microparticle, a nucleic acid is directly bound to the surface of the microparticle or is indirectly attached using a "bridge" or "spacer". The amino groups of the gelatin lysine groups are easily derivatized to provide sites for direct coupling of a compound. Alternatively, spacers (i.e., linking molecules and derivatizing moieties on targeting ligands) such as avidin-biotin are also useful to indirectly couple targeting ligands to the microparticles. Stability of the microparticle is controlled by the amount of glutaraldehyde-spacer crosslinking induced by the EDC hydrochloride. A controlled release medium is also empirically determined by the final density of glutaraldehyde-spacer crosslinks.

In one embodiment, the present invention contemplates microparticles formed by spray-drying a composition comprising fibrinogen or thrombin with a nucleic acid. Preferably, these microparticles are soluble and the selected protein (i.e., fibrinogen or thrombin) creates the walls of the microparticles. Consequently, the nucleic acids are incorporated within, and between, the protein walls of the microparticle. Heath et al., Microparticles And Their Use In Wound Therapy. United States Patent No. 6,113,948 (herein incorporated by reference). Following the application of the microparticles to living tissue, the subsequent reaction between the fibrinogen and thrombin creates a tissue sealant thereby releasing the incorporated compound into the immediate surrounding area.

One having skill in the art will understand that the shape of the microspheres need not be exactly spherical; only as very small particles capable of being sprayed or spread into or onto a surgical site (i.e., either open or closed). In one embodiment, microparticles are comprised of a biocompatible and/or biodegradable material selected from the group consisting of polylactide, polyglycolide and copolymers of lactide/glycolide (PLGA), hyaluronic acid, modified polysaccharides and any other well known material.

Experimental

Example I

Construction Of All-In-One sgRNA-Nme1Cas9-AAV Vector Plasmid

Bacterial Nme1Cas9 gene has been codon-optimized for expression in humans, and cloned into an AAV2 plasmid under U1a ubiquitous promoter. Guide RNA is under U6 promoter. The *cas9* gene contains four nuclear localization signals and three HA tag sequences in tandem. Spacer sequences were inserted into the crRNA cassette by digesting the plasmid with SapI restriction enzyme using annealed synthetic oligonucleotides to generate a duplex with overhangs compatible with those generated by SapI digested backbone.

The human-codon optimized Nme1Cas9 gene under the control of the U1a promoter and a sgRNA cassette driven by the U6 promoter were cloned into an AAV2 plasmid backbone. The NmeCas9 ORF was flanked by four nuclear localization signals – two on each terminus – in addition to a triple-HA epitope tag. This plasmid is available through Addgene (plasmid ID 112139). See, Figure 64. Oligonucleotides with spacer sequences targeting *Hpd, Pcsk9,* and *Rosa26* were inserted into the sgRNA cassette by ligation into a SapI cloning site.

AAV vector production was performed at the Horae Gene Therapy Center at the University of Massachusetts Medical School. Briefly, plasmids were packaged in AAV8 capsids by triple-plasmid transfection in HEK293 cells and purified by sedimentation as previously described. Gao et al., "Introducing genes into mammalian cells: viral vectors" In: Green MR, Sambrook J, editors. Molecular cloning: a laboratory manual. Volume 2. 4th ed. New York: Cold Spring Harbor Laboratory Press; 2012. p. 1209–13. The off-target profiles of these spacers were predicted computationally using the Bioconductor package CRISPRseek. Search parameters were adapted to Nme1Cas9 settings: gRNA.size = 24, PAM= "NNNNGATT," PAM.- size = 8,

RNA.PAM.pattern = "NNNNGNNN$," weights = c(0, 0, 0, 0, 0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583), max.mismatch = 6, allowed.mismatch.PAM= 7, topN = 10,000, min.score = 0.

Example II

Cell Culture And Transfection

Mouse Hepa1-6 hepatoma cells were cultured in DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37°C incubator with 5% CO_2. Human HEK293T cells and PLB985 cells were cultured in DMEM and RPMI media respectively. Both were supplemented with 10% FBS and 1% Penicillin/Streptomycin (Gibco). Transient transfections of Hepa 1-6 cells were performed using Lipofectamine LTX whereas Polyfect transfection reagent (Qiagen) was used for HEK293T cells. For transient transfection, approximately $1x10^5$ cells per well were cultured in 24-well plate 24 hours before transfection. Each well was transfected with 500ng all-in-one sgRNA-Nme1Cas9-AAV plasmids, using Lipofectamine LTX with Plus Reagent (ThermoFisher) according to the manufacturer's protocol. HEK293T cells were transfected with 400 ng of all-in-one plasmid expressing Nme1Cas9 and sgRNA in 24-well plate according to manufacturer's guidelines (e.g., *Psck9* & *Rosa26*).

All cell lines were maintained in a 37°C incubator with 5% CO_2. Mouse Hepa1-6 hepatoma and HEK293T cells were cultured in DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco). K562 cells were grown in the same conditions but using IMDM. IMR-90 cells were cultured in EMEM and10% FBS. Finally, HDFa cells were grown in DMEM and 20% FBS.

Example III

Expression And Purification Of Nme1Cas9

Nme1Cas9 was cloned into a pMCSG7 vector containing a T7 promoter followed by 6X His-tag and then a tobacco etch virus (TEV) protease cleavage site. This construct was transformed into Rosetta2 DE3 strain of *E. coli* and Nme1Cas9 was expressed. Briefly, bacterial culture was grown at 37 °C until OD600 of 0.6 was reached. At this point the temperature was lowered to 18 °C followed by addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were grown overnight, and then harvested for purification.

Purification of Nme1Cas9 was performed in three steps: Nickel affinity chromatography, cation exchange chromatography, and then size exclusion chromatography. The detailed protocols for these can be found in previous publications (Jinek *et al.*, *Science* 337, 816-821, 2012).

Example IV

Ribonucleoprotein (RNP) Delivery Of Nme1Cas9

RNP delivery of Nme1Cas9 was performed using the Neon transfection system (ThermoFisher). Approximately 20 picomoles of Nme1Cas9 and 25 picomoles of sgRNA were mixed in buffer R and incubated at room temperature for 20-30 minutes. This preassembled complex was then mixed with 50,000-100,000 cells, and electroporated using 10 µL Neon tips. After electroporation, cells were plated in 24-well plates containing the appropriate culture media without antibiotics.

Example V

DNA isolation from cells and tissue

Genomic DNA was isolated 72 hours post-transfection from cells via DNeasy® Blood and Tissue kit (Qiagen) according to the manufacturer's protocol. Mice were sacrificed and liver tissue was harvested 10 days post-hydrodynamic injection or 50 days post-tail vein vector administration, and genomic DNA was isolated with a DNeasy® Blood and Tissue kit (Qiagen) according to the manufacturer's protocol.

Example VI

Indel Analysis

50ng of genomic DNA was used for PCR amplification with genomic site-specific primers and High Fidelity® 2X PCR Master Mix (New England Biolabs). For TIDE analysis, 30µl of PCR product was purified using QIAquick® PCR Purification Kit (Qiagen), and subjected to Sanger sequencing. Indel values were obtained using the TIDE web tool (*tide-calculator.nki.nl/*) as described previously. Brinkman *et al.*, *Nucl. Acids Res.* (2014).

For the T7 Endonuclease I (T7EI) assay, 10µl of the PCR product was hybridized and treated with 0.5µl T7 Endonuclease I (New England Biolabs) in 1X NEB Buffer 2 for 1 hour.

The samples were run on a 2.5% agarose gel and quantified with ImageMaster-TotalLab® program. Indel percentages were calculated as previously described. Guschin *et al.*, Engineered Zinc Finger Proteins: Methods and Protocols (2010).

Example VII

GUIDE-Seq For Off-Target Analysis

GUIDE-seq analysis was performed as previously described. Tsai *et al.*, *Nature Biotechnology* (2014), Bolukbasi *et al.*, *Nature Methods* (2015a); Amrani et al., biorxiv.org/content/early/2017/08/04/172650 (2017).

Briefly, Hepa1-6 cells were transfected with 500ng of all-in-one sgRNA-Nme1Cas9-AAV plasmids and 7.5 pmol of annealed GUIDE-seq oligonucleotide using Lipofectamine LTX® with Plus® Reagent (ThermoFisher), for the two spacers targeting *Pcsk9* and *Rosa26* genes. Genomic DNA was extracted with a DNeasy® Blood and Tissue kit (Qiagen) at 72 hours after transfection following the manufacturer protocol. Library preparations, deep sequencing, and reads analysis were performed as previously described. Tsai *et al.*, *Nature Biotechnology* (2014), Bolukbasi *et al.*, *Nature Methods* (2015a); Amrani et al., *biorxiv.org/content/early/2017/08/04/172650* (2017).

Example IX

AAV Vector Production

Plasmids were packaged in AAV8 by triple-plasmid transfection in HEK 293 cells and purified by sedimentation as previously described at the Horae Gene Therapy Center at the University of Massachusetts Medical School. Gao GP, Sena-Esteves M. Introducing Genes into Mammalian Cells: Viral Vectors. In: Green MR, Sambrook J, eds. Molecular Cloning, Volume 2: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press; 2012:1209–1313.

Example X

Animals, AAV Vector Injections, And Liver Tissue Processing

All animal experiments were approved under the guidelines of the University of Massachusetts Medical School Institutional Animal Care and Use Committee. For hydrodynamic injections, 2.5mL of 30 μg of endotoxin-free sgRNA-Nme1Cas9-AAV plasmid targeting *Pcsk9*,

or PBS as a control, were injected via tail vein into 9-18 weeks old female C57BL/6 mice. For the AAV8 vector injections, 9-18 weeks old female C57BL/6 mice were injected with $4x10^{11}$ genome copies per mouse via tail vein. 8-week-old female C57BL/6NJ mice were used for genome editing experiments in vivo. For ex vivo experiments, embryos that were advanced to two-cell stage were transferred into the oviduct of E0.5 pseudo-pregnant female mice.

Mice were euthanized by CO_2 and liver was collected. Tissues were fixed in 4% paraformaldehyde overnight, and embedded in paraffin, sectioned and stained with hematoxylin and eosin (H&E).

Example XI

Serum Analysis

Blood (~ 200 µL) was drawn from the facial vein at 0, 25, and 50 days post vector administration. Serum was isolated using a serum separator (BD, Cat. No. 365967) and stored under – 80 °C until assay. Serum cholesterol levels were measured by Infinity™ colorimetric endpoint assay (Thermo-Scientific) following the manufacturer's protocol. Briefly, serial dilutions of Data-Cal™ Chemistry Calibrator were prepared in PBS. In a 96-well plate, 2 µL of mice sera or calibrator dilution was mixed with 200 µL of Infinity™ cholesterol liquid reagent, then incubated at 37 °C for 5 min. The absorbance was measured at 500 nm using a BioTek Synergy® HT microplate reader.

Example XII

Discovery Of Cas9 Orthologs With Hyper-Evolved PIDs

Nme1Cas9 sequence was blasted to find all Cas9 orthologs in *Neisseria* species. Orthologs with >80% identity to Nme1Cas9 were selected for the remainder of this analysis. The PIDs of each was then aligned using ClustalW2 with that of Nme1Cas9 (from 820^{th} amino acid to 1082^{nd}) and those with clusters of mutations in the PID were selected.

Nme1Cas9 peptide sequence was used as a query in BLAST searches to find all Cas9 orthologs in Neisseria meningitidis strains. Orthologs with >80% identity to Nme1Cas9 were selected for 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 (tree.bio.ed.ac.uk/software/figtree).

Example XIII

Cloning And Purification Of Nme2 and Nme3 Cas9 And Acr Orthologs

The PIDs of Nme2Cas9 and Nme3Cas9 were ordered as gBlocks (IDT) to replace the PID of Nme1Cas9 using Gibson Assembly (NEB) in a bacterial expression plasmid pMSCG7 with 6X His-tag. The construct was transformed into *E. coli*, expressed and purified as previously described.

Briefly, Rosetta (DE3) cells containing the respective Cas9 plasmids were grown at 37°C to an optical density 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 pH 7.5, 500 mM NaCl, 5 mM imidazole, 1 mM DTT) supplemented with Lysozyme and protease inhibitor cocktail (Sigma).

The lysate was then run through a Ni-NTA agarose column (Qiagen), the bound protein was eluted with 300mM imidazole and dialyzed into storage buffer (20 mM HEPES pH 7.5, 250 mM NaCl, 1 mM DTT). For Acr proteins, 6x His tagged proteins were expressed in *E. coli* strain BL21 Rosetta (DE3). Cells were grown at 37° C to an optical density ($OD_{600\ nm}$) 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 (50 mM Tris pH 7.5, 500 mM NaCl, 5 mM imidazole, 1 mM DTT) supplemented with 1 mg/mL Lysozyme and protease inhibitor cocktail (Sigma) and protein was purified using the same protocol as for Cas9. The 6X His tag was removed by incubation with Tobacco Etch Virus (TEV) protease overnight at 4 °C to isolate successfully cleaved, untagged Acrs.

Example IVX

In vitro PAM Discovery Assay

A library of protospacers with randomized PAM sequences was generated using overlapping PCRs, with the forward primer containing the 10-nucleotide randomized PAM.

The library was gel purified and subjected to *in vitro* cleavage reaction by purified Cas9 along with *in vitro* transcribed sgRNAs. 300 nM Cas9:sgRNA complex was used to cleave 300 nM of the target fragment in 1X NE Buffer 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 gel with 1X TAE. The cleavage product was purified and subjected to library preparation. The library was sequenced using the Illumina NextSeq500® sequencing platform and analyzed. Sequence logos were generated using R.

Example XV

Transfections And Mammalian Genome Editing

Humanized Nme2Cas9 was cloned into pCDest2 plasmid previously used for Nme1Cas9 and SpyCas9 expression using Gibson Assembly. Transfection of HEK293T and HEK293T-TLR 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 approximately $1x10^5$ cells per well that had been cultured in 24-well plates 24 hours before transfection. For K562 cells stably expressing Nme2Cas9, 50,000 – 150,000 cells were electroporated with 500 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).

Example XVI

Lentiviral transduction of K562 cells to stably express Nme2Cas9

K562 cells stably expressing Nme2Cas9 were generated as previously described. 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) as recommended by the manufacturer. After 24 hours, the medium was aspirated from the transfected cells and replaced with fresh 1 mL of fresh DMEM media.

The next day, the supernatant containing the virus from the transfected cells was collected and filtered through 0.45 µm filter. 10 uL of the undiluted supernatant along with 2.5

μg of Polybrene was used to transduce ~ 1 million K562 cells in 6-well plates. The transduced cells were selected using 2.5 μg/mL of Puromycin containing media.

Example XVII

RNP Delivery For Mammalian Genome Editing

For RNP experiments, a Neon electroporation system was used. 40 picomoles of 3X NLS Nme2Cas9 along with 50 picomoles of *in vitro* 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 XVIII

GUIDE-Seq

GUIDE-Seq experiments were performed as described previously with minor modifications (Amrani et al., 2018).

Briefly, HEK293T cells were transfected with 200 ng of Cas9, 200 ng of sgRNA, and 7.5 pmol of annealed GUIDE-seq oligonucleotide using Polyfect (Qiagen) for guides targeting dual sites with SpyCas9 or Nme2Cas9. 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 protocol. Library preparation and sequencing were performed exactly as described previously.

For analysis, sites that matched a sequence with ten mismatches with the target site were considered potential off-target sites. Data were analyzed using the Bioconductor package GUIDEseq version 1.1.17 (Zhu et al., 2017).

Example XIX

Targeted Deep Sequencing And Analysis

Targeted deep sequencing was used to confirm the results of GUIDE-Seq and more quantitatively measure indel rates. A two-step PCR amplification was used to produce DNA

fragments for each on- and off-target site. For SpyCas9, the top off-target locations were selected.

In the first step, locus-specific primers bearing universal overhangs with complementary ends to the adapters were mixed with 2X Phusion® PCR master mix (NEB) to generate fragments bearing the overhangs. In the second step, the purified PCR products were amplified with a universal forward primer and and indexed reverse primers.

Full-size products (~250bp in length) were gel-extracted and sequenced using a paired-end MiSeq run. MiSeq data analysis was performed exactly as previously described (Amrani 2018).

Example XX

Off-Target Analysis Using CRISPRseek

Global off-target analyses for TS25 and TS47 were performed using the Bioconductor package CRISPRseek.

Minor changes were made to accommodate for characteristics of Nme2Cas9 not shared with SpyCas9. Specifically, the following changes were used: gRNA.size = 24, PAM = "NNNNCC", PAM.size = 6, RNA.PAM.pattern = "NNNNCN", off-target sites with less than 6 mismatches were collected. The top potential off-target sites based on the number and position of mismatches were selected. gDNA from cells targeted by each respective sgRNA was used to amplify each off-target locus and analyzed by TIDE.

Example XXI

In vivo AAV8.Nme2Cas9 Delivery And Liver Tissue Processing

All animal procedures were reviewed and approved by The Institutional Animal Care and Use Committee (IACUC) at University of Massachusetts Medical School.

For the AAV8 vector injections, 8 weeks old female C57BL/6 mice were injected with 4 $x10^{11}$ genome copies per mouse via tail vein targeting *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, and embedded in paraffin, sectioned and stained with hematoxylin and eosin (H&E). Blood was drawn from 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 manufacturer's protocol and as previously described (Ibraheim et al, 2018).

Example XXII

Animals and liver tissue processing

For hydrodynamic injections, 2.5 mL of 30 μg of endotoxin-free AAV-sgRNA-hNme1Cas9 plasmid targeting *Pcsk9* or 2.5 mL PBS was injected by tail vein into 9- to 18-week-old female C57BL/6 mice. Mice were euthanized 10 days later and liver tissue was harvested. For the AAV8 vector injections, 12- to 16-week-old female C57BL/6 mice were injected with 4 × 10^{11} genome copies per mouse via tail vein, using vectors targeting *Pcsk9* or *Rosa26*. Mice were sacrificed 14 and 50 days after vector administration and liver tissues were collected for analysis.

For *Hpd* targeting, 2 mL PBS or 2 mL of 30 μg of endotoxin-free AAV-sgRNA-hNme1Cas9 plasmid was administered into 15- to 21-week-old Type 1 Tyrosinemia *Fah* knockout mice (Fahneo) via tail vein. The encoded sgRNAs targeted sites in exon 8 (sgHpd1) or exon 11 (sgHpd2). The HT1 mice were fed with 10 mg/L NTBC (2-(2-nitro-4--trifluoromethylbenzoyl)-1,3-cyclohexanedione) (Sigma-Aldrich, Cat. No. PHR1731-1G) in drinking water when indicated. Both sexes were used in these experiments. Mice were maintained on NTBC water for seven days post injection and then switched to normal water. Body weight was monitored every 1–3 days. The PBS-injected control mice were sacrificed when they became moribund after losing 20% of their body weight after removal from NTBC treatment.

Mice were euthanized according to our protocol and liver tissue was sliced and fragments stored at − 80 °C. Some liver tissues were fixed in 4% formalin overnight, embedded in paraffin, sectioned and stained with hematoxylin and eosin (H&E).

XXIII

Western Blot

Liver tissue fractions were ground and resuspended in 150 μL of RIPA lysis buffer. Total protein content was estimated by Pierce™ BCA Protein Assay Kit (Thermo-- Scientific)

following the manufacturer's protocol. A total of 20 μg of protein from tissue or 2 ng of Recombinant Mouse Proprotein Convertase 9/PCSK9 Protein (R&D Systems, 9258-SE-020) were loaded onto a 4–20% Mini-- Rotean® TGX™ Precast Gel (Bio-Rad). The separated bands were transferred onto PVDF membrane and blocked with 5% Blocking-Grade Blocker solution (Bio-Rad) for 2 h at room temperature. Membranes were incubated with rabbit anti-GAPDH (Abcam ab9485, 1:2000) or goat anti-PCSK9 (R&D Systems AF3985, 1:400) antibodies overnight at 4 °C. Membranes were washed five times in TBST and incubated with horseradish peroxidase (HRP)-conjugated goat anti-rabbit (Bio-Rad 1,706,515, 1:4000) and donkey anti-goat (R&D Systems HAF109, 1:2000) secondary antibodies for 2 h at room temperature. The membranes were washed five times in TBST and visualized with Clarity™ western ECL substrate (Bio-Rad) using an M35A X-OMAT Processor (Kodak).

Example XXIV

Humoral Immune Response

Humoral IgG1 immune response to Nme1Cas9 was measured by ELISA (Bethyl; Mouse IgG1 ELISA Kit, E99-- 105) following manufacturer's protocol with a few modifications. Briefly, expression and three-step purification of Nme1Cas9 and SpyCas9 was performed. A total of 0.5 μg of recombinant Nme1Cas9 or SpyCas9 proteins suspended in 1X coating buffer (Bethyl) were used to coat 96-well plates (Corning) and incubated for 12 h at 4 °C with shaking. The wells were washed three times while shaking for 5 min using 1X Wash Buffer. Plates were blocked with 1X BSA Blocking Solution (Bethyl) for 2 h at room temperature, then washed three times. Serum samples were diluted 1:40 using PBS and added to each well in duplicate. After incubating the samples at 4° C for 5 h, the plates were washed 3 times for 5 min and 100 μL of biotinylated anti-mouse IgG1 antibody (Bethyl; 1: 100,000 in 1 x BSA Blocking Solution) was added to each well. After incubating for 1 h at room temperature, the plates were washed four times and 100 μL of TMB Substrate was added to each well. The plates were allowed to develop in the dark for 20 min at room temperature and 100 μL of ELISA Stop Solution was then added per well. Following the development of the yellow solution, absorbance was recorded at 450 nm using a BioTek Synergy® HT microplate reader.

Example XXV

Zygote Incubation And Transfection

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.

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 4×10^{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., " All-in-One Adeno-associated Virus Delivery and Genome Editing by Neisseria meningitidis Cas9 in vivo" *Genome Biology* 19:137 (2018).

For an 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.sgTyr 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.

Example XXVI

Quantification And Statistical Analyses

An analysis of in vitro PAM discovery data was performed using R. GraphPad Prism 6® for all statistical analyses. For mammalian cell experiments using Nme2Cas9, 3 independent replicates were performed and indel percentages were calculated using TIDE software, with error bars depicting s.e.m. The TIDE parameters were set to quantify indels <20 nucleotides for all figures. For side-by-side comparisons of Nme2Cas9 and SpyCas9, average indel percentages were calculated using Microsoft Excel. For in vivo experiments in mice, n = 5 for control and test subjects. P values were calculated by unpaired two-tailed t-test.

SEQUENCE LISTING

<110> University of Massachusetts

<120> Targeted Crispr Delivery Platforms

<130> P000818SG2

<150> PCT/US2018/060126
<151> 2018-11-09

<150> US 62/584,310
<151> 2017-11-10

<150> US 62/596,375
<151> 2017-12-08

<150> US 62/667,084
<151> 2018-05-04

<160> 314

<170> PatentIn version 3.5

<210> 1
<211> 145
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 1
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccuuuc ucauuucgga aacgaaauga 60

gaaccguugc uacaauaagg ccgucugaaa agaugugccg caacgcucug ccccuuaaag 120

cuucugcuuu aaggggcauc guuua 145

<210> 2
<211> 121
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 2

nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuaaagcuuc ugcuuuaagg ggcaucguuu 120

a 121

<210> 3
<211> 111
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 3
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuuucuaagg ggcaucguuu a 111

<210> 4
<211> 105
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 4
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuuucuaagg ggcau 105

<210> 5
<211> 99
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 5
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcuuc ugcaucguu 99

<210> 6
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(23)
<223> n is a, c, g, t or u

<400> 6
nnnnnnnnnn nnnnnnnnnn nnnguuguag cucccgaaac guugcuacaa uaaggccguc 60

ugaaaagaug ugccgcaacg cucugcuucu gcaucguuua 100

<210> 7
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(21)
<223> n is a, c, g, t or u

<400> 7
nnnnnnnnnn nnnnnnnnnn nguuguagcu cccgaaacgu ugcuacaaua aggccgucug 60

aaaagaugug ccgcaacgcu cugcccuucu gggcaucguu 100

<210> 8
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 8

nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuucuagggg caucguu 107

<210> 9
<211> 105
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 9
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uucuggggca ucguu 105

<210> 10
<211> 103
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 10
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccu ucugggcauc guu 103

<210> 11
<211> 101
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 11
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugccuu cuggcaucgu u 101

<210> 12
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(23)
<223> n is a, c, g, t or u

<400> 12
nnnnnnnnnn nnnnnnnnnn nnnguuguag cucccgaaac guugcuacaa uaaggccguc 60

ugaaaagaug ugccgcaacg cucugccuuc uggcaucguu 100

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 13
gagggagaga ggugagcgga ugaa 24

<210> 14
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 14
ggggagagag gugagcggau gaa 23

<210> 15
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 15
guucuccaag cccucggacc ucgu 24

<210> 16

<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 16
gucuccaagc ccucggaccu cgu 23

<210> 17
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 17
gcuggauuac uguguggguag aggg 24

<210> 18
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 18
guggauuacu guguggguaga ggg 23

<210> 19
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 19
agcttgagca aagggagaga ggtgagcgga tgaagggaga ttggtgagta tc 52

<210> 20
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 20
cgcttcgcgg cttctccaag ccctcggacc tcgtgggcgt cttctcctgc gt 52

<210> 21
<211> 52

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 21
gaattcacta gctggattac tgtgtggtag agggaggtga ttagcacctg tg 52

<210> 22
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 22
gaatatcagg agactaggaa ggag 24

<210> 23
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 23
gaggccta 8

<210> 24
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 24
ggacaggagt cgccagaggc cggt 24

<210> 25
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 25
ggtggattt 9

<210> 26
<211> 23
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 26
gcacctgcct cgtggaatac ggt 23

<210> 27
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 27
aaacctac 8

<210> 28
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 28
atgtggctct ggttctgggt acttttatct gtcccctcca ccccacagtg gg 52

<210> 29
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 29
cagataagga atctgcctaa caggaggtgg gggttagacg aatatcagga ga 52

<210> 30
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 30
ggggttagac gaatatcagg agactaggaa ggaggaggcc taaggatggg gg 52

<210> 31
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 31
cccacccggc ggcgcctccc tgcagggctg ctccccagcc caaaccgccg cg 52

<210> 32
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 32
tccgagagct cagctagtct tcttcctcca acccgggccc tatgtccact tc 52

<210> 33
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 33
tgggtacttt tatctgtccc ctccacccca cagtggggcc actagggaca gg 52

<210> 34
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 34
gtaggggagc tgcccaaatg aaaggagtga gaggtgaccc gaatccacag ga 52

<210> 35
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 35
tagcacctct ccatcctctt gctttctttg cctggacacc ccgttctcct gt 52

<210> 36
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 36
gtctcccttg cgtcccgcct ccccttcttg taggcctgca tcatcaccgt tt 52

<210> 37
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 37
cctcacccaa ccccatgccg tgttcactcg ctgggttccc ttttccttct cct 53

<210> 38
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 38
gcgcaggaca ggagtcgcca gaggccggtg gtggatttcc tccccgcatc tc 52

<210> 39
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 39
cgcggggacg cccagcggcc ggatatcagc tgccacgccc gcgtgggcgg a 51

<210> 40
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 40
gattccaata gatctgtgtg tccctctccc cacccgtccc tgtccggctc tc 52

<210> 41
<211> 52
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 41
tgacccctgg ccttctcccc gctccaacgc cctcaacccc acacgcacac ac 52

<210> 42
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 42
tccctctccc cacccgtccc tgtccggctc tccgccttcc cctgcccccc tc 52

<210> 43
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 43
acacgcacac actcactcac ccacacagac acacacgtcc tcactctcga ag 52

<210> 44
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 44
taagcacagt ggaagaattt cattctgttc tcagttttcc tggattatgc ct 52

<210> 45
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 45
ttcattctgt tctcagtttt cctggattat gcctggcacc attaaagaaa at 52

<210> 46
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 46
ggttctgggt acttttatct gtcc 24

<210> 47
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 47
cctccacc 8

<210> 48
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 48
tggcttagca cctctccat 19

<210> 49
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 49
agaactcagg accaacttat tctg 24

<210> 50
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 50
gtctgcctaa caggaggtgg gggt 24

<210> 51
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 51
tagacgaa 8

<210> 52
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 52
tggcttagca cctctccat 19

<210> 53
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 53
agaactcagg accaacttat tctg 24

<210> 54
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 54
gaatatcagg agactaggaa ggag 24

<210> 55
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 55
gaggccta 8

<210> 56
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 56

tggcttagca cctctccat 19

<210> 57
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 57
agaactcagg accaacttat tctg 24

<210> 58
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 58
gcctccctgc agggctgctc cc 22

<210> 59
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 59
cagcccaa 8

<210> 60
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 60
agaggagcct tctgactgct gcaga 25

<210> 61
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 61
atgacagaca caaccagagg gca 23

<210> 62
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 62
gagctagtct tcttcctcca accc 24

<210> 63
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 63
gggcccta 8

<210> 64
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 64
tggcttagca cctctccat 19

<210> 65
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 65
agaactcagg accaacttat tctg 24

<210> 66
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 66
gatctgtccc ctccacccca cagt 24

<210> 67
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 67
ggggccac 8

<210> 68
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 68
tggcttagca cctctccat 19

<210> 69
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 69
agaactcagg accaacttat tctg 24

<210> 70
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 70
ggcccaaatg aaaggagtga gagg 24

<210> 71
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 71
tgacccga 8

<210> 72
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 72
tccgtcttcc tccactcc 18

<210> 73
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 73
taggaaggag gaggcctaag 20

<210> 74
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 74
gcatcctctt gctttctttg cctg 24

<210> 75
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 75
gacacccc 8

<210> 76
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 76
tccgtcttcc tccactcc 18

<210> 77

<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 77
taggaaggag gaggcctaag 20

<210> 78
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 78
ggagtcgcca gaggccggtg gtgg 24

<210> 79
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 79
atttcctc 8

<210> 80
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 80
agaggagcct tctgactgct gcaga 25

<210> 81
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 81
atgacagaca caaccagagg gca 23

<210> 82
<211> 24

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 82
gcccagcggc cggatatcag ctgc 24

<210> 83
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 83
cacgcccg 8

<210> 84
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 84
agaggagcct tctgactgct gcaga 25

<210> 85
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 85
atgacagaca caaccagagg gca 23

<210> 86
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 86
ggaagggaac atattactat tgc 23

<210> 87
<211> 8
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 87
tttccctc 8

<210> 88
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 88
tagagaactg ggtagtgtg 19

<210> 89
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 89
ccaatattgc atgggatgg 19

<210> 90
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 90
gtggagtggc ctgctatcag ctac 24

<210> 91
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 91
ctatccaa 8

<210> 92
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 92
tagagaactg ggtagtgtg 19

<210> 93
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 93
ccaatattgc atgggatgg 19

<210> 94
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 94
gaggaaggga acatattact attg 24

<210> 95
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 95
ctttccct 8

<210> 96
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 96
tagagaactg ggtagtgtg 19

<210> 97
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 97
ccaatattgc atgggatgg 19

<210> 98
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 98
gtgaattctc atcagctaaa atgc 24

<210> 99
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 99
caagcctt 8

<210> 100
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 100
tagagaactg ggtagtgtg 19

<210> 101
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 101
ccaatattgc atgggatgg 19

<210> 102
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 102
gctcactcac ccacacagac acac 24

<210> 103
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 103
acgtcctc 8

<210> 104
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 104
gtacatgaag caactccagt ccca 24

<210> 105
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 105
atcaaattcc agcaccgagc gc 22

<210> 106
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 106
ggaagaattt cattctgttc tcag 24

<210> 107
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 107
ttttcctg 8

<210> 108
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 108
tggtgattat gggagaactg gagc 24

<210> 109
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 109
accattgagg acgtttgtct cac 23

<210> 110
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 110
gctcagtttt cctggattat gcct 24

<210> 111
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 111
ggcaccat 8

<210> 112
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 112
tggtgattat gggagaactg gagc 24

<210> 113
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 113
accattgagg acgtttgtct cac 23

<210> 114
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 114
gcgttggagc ggggagaagg ccag 24

<210> 115
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 115
gggtcact 8

<210> 116
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 116
gtacatgaag caactccagt ccca 24

<210> 117
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 117

atcaaattcc agcaccgagc gc 22

<210> 118
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 118
gggccgcgga gatagctgca gggc 24

<210> 119
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 119
ggggcccc 8

<210> 120
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 120
agaggagcct tctgactgct gcaga 25

<210> 121
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 121
atgacagaca caaccagagg gca 23

<210> 122
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 122
gcccacccgg cggcgcctcc ctgc 24

<210> 123
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 123
agggctgc 8

<210> 124
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 124
agaggagcct tctgactgct gcaga 25

<210> 125
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 125
atgacagaca caaccagagg gca 23

<210> 126
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 126
gcgtggcagc tgatatccgg ccgc 24

<210> 127
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 127
tgggcgtc 8

<210> 128
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 128
agaggagcct tctgactgct gcaga 25

<210> 129
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 129
atgacagaca caaccagagg gca 23

<210> 130
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 130
gccgcggcgc gacgtggagc cagc 24

<210> 131
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 131
cccgcaaa 8

<210> 132
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 132
agaggagcct tctgactgct gcaga 25

<210> 133
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 133
atgacagaca caaccagagg gca 23

<210> 134
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 134
gtgctcccca gcccaaaccg ccgc 24

<210> 135
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 135
ggcgcgac 8

<210> 136
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 136
agaggagcct tctgactgct gcaga 25

<210> 137
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 137
atgacagaca caaccagagg gca 23

<210> 138

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 138
gtcagattgg cttgctcgga attg 24

<210> 139
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 139
ccagccaa 8

<210> 140
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 140
ggcataagga aatcgaaggt c 21

<210> 141
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 141
catgtcctca agtcaagaac aag 23

<210> 142
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 142
gctgggtgaa tggagcgagc agcg 24

<210> 143
<211> 8

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 143
tcttcgag 8

<210> 144
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 144
gtacatgaag caactccagt ccca 24

<210> 145
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 145
atcaaattcc agcaccgagc gc 22

<210> 146
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 146
gtcctggagt gacccctggc cttc 24

<210> 147
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 147
tccccgct 8

<210> 148
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 148
gtacatgaag caactccagt ccca 24

<210> 149
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 149
atcaaattcc agcaccgagc gc 22

<210> 150
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 150
gatcctggag tgacccctgg cctt 24

<210> 151
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 151
ctccccgc 8

<210> 152
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 152
gtacatgaag caactccagt ccca 24

<210> 153
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 153
atcaaattcc agcaccgagc gc 22

<210> 154
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 154
gtgtgtccct ctccccaccc gtcc 24

<210> 155
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 155
ctgtccgg 8

<210> 156
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 156
gtacatgaag caactccagt ccca 24

<210> 157
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 157
atcaaattcc agcaccgagc gc 22

<210> 158
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 158
gttggagcgg ggagaaggcc aggg 24

<210> 159
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 159
gtcactcc 8

<210> 160
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 160
gtacatgaag caactccagt ccca 24

<210> 161
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 161
atcaaattcc agcaccgagc gc 22

<210> 162
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 162
gcgttggagc ggggagaagg ccag 24

<210> 163
<211> 8
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic

<400> 163
gggtcact 8

<210> 164
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 164
gtacatgaag caactccagt ccca 24

<210> 165
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 165
atcaaattcc agcaccgagc gc 22

<210> 166
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 166
gtaccctcca ataatttggc tggc 24

<210> 167
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 167
aattccga 8

<210> 168
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 168
ggcataagga aatcgaaggt c 21

<210> 169
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 169
catgtcctca agtcaagaac aag 23

<210> 170
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 170
gataatttgg ctggcaattc cgag 24

<210> 171
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 171
caagccaa 8

<210> 172
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 172
ggcataagga aatcgaaggt c 21

<210> 173
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 173
catgtcctca agtcaagaac aag 23

<210> 174
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 174
gcaggggcca ggtgtccttc tctg 24

<210> 175
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 175
ggggcctc 8

<210> 176
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 176
acacgggcag catgggaata gtc 23

<210> 177
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 177
gctaggggag agtcccactg tcca 24

<210> 178
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 178

gaatggcagg cggaggttgt actg 24

<210> 179
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 179
ggggccag 8

<210> 180
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 180
cctgtgtggc tttgctttgg tcg 23

<210> 181
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 181
gtagggtgtg atgggaggct aagc 24

<210> 182
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 182
gagtgagaga gtgagagaga gaca 24

<210> 183
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 183
cgggccag 8

<210> 184
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 184
cctgtgtggc tttgctttgg tcg 23

<210> 185
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 185
gtagggtgtg atgggaggct aagc 24

<210> 186
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 186
gtgagcaggc acctgtgcca acat 24

<210> 187
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 187
gggcccgc 8

<210> 188
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 188
cctgtgtggc tttgctttgg tcg 23

<210> 189
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 189
gtagggtgtg atgggaggct aagc 24

<210> 190
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 190
gcgtgggggc tccgtgcccc acgc 24

<210> 191
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 191
gggtccat 8

<210> 192
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 192
ggaggaagag tagctcgccg agg 23

<210> 193
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 193
agaccgagtg gcagtgacag caag 24

<210> 194
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 194
gcatgggcag gggctggggt gcac 24

<210> 195
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 195
aggcccag 8

<210> 196
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 196
agggagaggg aagtgtgggg aagg 24

<210> 197
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 197
gtcttcctgc tctgtgcgca cgac 24

<210> 198
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 198
gaaaattgtg atttccagat ccac 24

<210> 199

<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 199
aagcccaa 8

<210> 200
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 200
gttgggggct ctaagttatg tat 23

<210> 201
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 201
cttcatctgt atcttcagga tca 23

<210> 202
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 202
gagcagaaaa aattgtgatt tcc 23

<210> 203
<211> 8
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 203
agatccac 8

<210> 204
<211> 23

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 204
gttgggggct ctaagttatg tat 23

<210> 205
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 205
cttcatctgt atcttcagga tca 23

<210> 206
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (2)..(4)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (6)..(25)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (28)..(28)
<223> n is a, c, g, or t

<400> 206
gnnngnnnnn nnnnnnnnnn nnnnnggncc 30

<210> 207
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 207
guuguagcuc cccuuucuca uuucggaaac gaaaugagaa ccguugcuac aauaaggccg 60

ucugaaaaga ugugccgcaa cgcucugccc cuuaaagcuu cugcuuuaag gggcaucguu 120

ua 122

<210> 208
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 208
gagggagaga ggugagcgga ugaaguugua gcucccuuuc ucauuucgga aacgaaauga 60

gaaccguugc uacaauaagg cgcucugaaa agaugugccg caacgcucug cccuuaaagc 120

uucugcuuua aggggcaucg uuua 144

<210> 209
<211> 121
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 209
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuaaagcuuc ugcuuuaagg ggcaucguuu 120

a 121

<210> 210
<211> 111
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 210
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuuucuaagg ggcaucguuu a 111

<210> 211
<211> 105
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 211
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuuucuaagg ggcau 105

<210> 212
<211> 99
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 212
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcuuc ugcaucguu 99

<210> 213
<211> 99
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 213
gggagagagg ugagcggaug aaguuguagc ucccgaaacg uugcuacaau aaggccgucu 60

gaaaagaugu gccgcaacgc ucugcuucug caucguuua 99

<210> 214
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 214
ggagagaggu gagcggauga aguuguagcu cccgaaacgu ugcuacaaua aggccgucug 60

aaaagaugug ccgcaacgcu cugcccuucu gggcaucguu 100

<210> 215
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 215
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuucuagggg caucguu 107

<210> 216

<211> 105
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 216
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uucuggggca ucguu 105

<210> 217
<211> 96
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 217
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccu ucuggg 96

<210> 218
<211> 101
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 218
gagggagaga ggugagcgga ugaaguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugccuu cuggcaucgu u 101

<210> 219
<211> 145
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 219
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccuuuc ucauuucgga aacgaaauga 60

gaaccguugc uacaauaagg ccgucugaaa agaugugccg caacgcucug ccccuuaaag 120

cuucugcuuu aaggggcauc guuua 145

<210> 220
<211> 101
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 220
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccgaaa cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugccuu cuggcaucgu u 101

<210> 221
<211> 100
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(23)
<223> n is a, c, g, t or u

<400> 221
nnnnnnnnnn nnnnnnnnnn nnnguuguag cucccgaaac guugcuacaa uaaggccguc 60

ugaaaagaug ugccgcaacg cucugccuuc uggcaucguu 100

<210> 222
<211> 121
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 222
guuguagcuc ccuuucucau uucggaaacg aaaugagaac cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuaaagcuuc ugcuuuaagg ggcaucguuu 120

a 121

<210> 223
<211> 77
<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic

<400> 223
guuguagcuc ccgaaacguu gcuacaauaa ggccgucuga aaagaugugc cgcaacgcuc 60

ugccuucugg caucguu 77

<210> 224
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 224
guuguagcuc ccgaaacguu gcuacaauaa ggccguugaa aagugugccg caacgcucug 60

ccuucuggca ucguuua 77

<210> 225
<211> 121
<212> DNA
<213> N. meningitidis

<400> 225
guuguagcuc ccuuucucau uucggaaacg aaaugagaac cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuaaagcuuc ugcuuuaagg ggcaucguuu 120

a 121

<210> 226
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 226
guuguagcuc ccgaaacguu gcuacaauaa ggccgucuga aaagaugugc cgcaacgcuc 60

ugccuucugg caucguuua 79

<210> 227
<211> 121
<212> DNA
<213> N. meningitidis

<400> 227
guuguagcuc ccuuucucau uucggaaacg aaaugagaac cguugcuaca auaaggccgu 60

cugaaaagau gugccgcaac gcucugcccc uuaaagcuuc ugcuuuaagg ggcaucguuu 120

a 121

<210> 228
<211> 97
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 228
guuguagcuc ccgaaacguu gcuacaauaa ggccgucuga aaagaugugc cgcaacgcuc 60

ugccccuuaa agcuucugcu uuaaggggca ucguuua 97

<210> 229
<211> 427
<212> DNA
<213> N. meningitidis

<400> 229
ggcaggaaga gggcctattt cccatgattc cttcatattt gcatatacga tacaaggctg 60

ttagagagat aattagaatt aatttgactg taaacacaaa gatattagta caaaatacgt 120

gacgtagaaa gtaataattt cttgggtagt ttgcagtttt aaaattatgt tttaaaatgg 180

actatcatat gcttaccgta acttgaaagt atttcgattt cttggcttta tatatcttgt 240

ggaaaggacg aaacaccgga agagcattcc agctcttcag ttgtagctcc ctttctcatt 300

tcggaaacga aatgagaacc gttgctacaa taaggccgtc tgaaaagatg tgccgcaacg 360

ctctgcccct taaagcttct gctttaaggg gcatcgttta ttttttgtt taaactctag 420

aggccgc 427

<210> 230
<211> 3963
<212> DNA
<213> N. meningitidis

<400> 230
acgtcgactc tagaatggag gcggtactat gtagatgaga attcaggagc aaactgggaa 60

aagcaactgc ttccaaatat ttgtgatttt tacagtgtag ttttggaaaa actcttagcc 120

taccaattct tctaagtgtt ttaaaatgtg ggagccagta cacatgaagt tatagagtgt 180

tttaatgagg cttaaatatt taccgtaact atgaaatgct acgcatatca tgctgttcag 240

gctccgtggc cacgcaactc atacttaagc agacagtggt tcaaagtttt tttcttccat 300

ttcaggtgtc gtgaacaccg ccaccatggt gcctaagaag aagagaaagg tggaagataa 360

acgcccagca gctacaaaga aggcaggtca agccaagaaa aagaaagcag cattcaagcc 420

aaactcaatc aattacatcc tgggactgga catcggcatc gcatccgtcg ggtgggctat 480

ggtcgaaatc gacgaggagg agaaccccat ccgcctgatc gatctgggcg tgcgcgtgtt 540

tgagagggca gaggtgccta agaccggcga cagcctggcc atggcacgga gactggcacg 600

ctccgtgagg cgcctgaccc ggagaagggc ccacagactg ctgaggacac gccggctgct 660

gaagagggag ggcgtgctgc aggccgccaa cttcgatgag aatggcctga tcaagtccct 720

gcccaatacc ccttggcagc tgagggcagc cgccctggac cgcaagctga cacctctgga 780

gtggtccgcc gtgctgctgc acctgatcaa gcaccggggc tacctgtctc agagaaagaa 840

cgagggcgag acagccgata aggagctggg cgccctgctg aagggagtgg caggaaatgc 900

acacgccctg cagaccggcg actttcgcac accagccgag ctggccctga acaagttcga 960

gaaggagagc ggccacatcc gcaatcagcg gtctgactat agccacacct tctcccggaa 1020

ggatctgcag gccgagctga tcctgctgtt tgagaagcag aaggagttcg gcaacccaca 1080

cgtgtctggc ggcctgaagg agggcatcga gacactgctg atgacacagc ggcccgccct 1140

gagcggcgac gcagtgcaga agatgctggg acactgcacc tttgagccag ccgagcccaa 1200

ggccgccaag aatacctaca cagccgagcg gttcatctgg ctgacaaagc tgaacaatct 1260

gaggatcctg gagcagggaa gcgagcgccc actgaccgac acagagaggg ccaccctgat 1320

ggatgagccc taccgcaagt ccaagctgac atatgcacag gcaaggaagc tgctgggcct 1380

ggaggacacc gccttcttta agggcctgag atacggcaag gataacgccg aggcctctac 1440

actgatggag atgaaggcct atcacgccat cagcagggcc ctggagaagg agggcctgaa 1500

ggacaagaag tccccactga atctgtctcc cgagctgcag gatgagatcg gcaccgcctt 1560

tagcctgttc aagaccgacg aggatatcac aggcagactg aaggacagga tccagccaga 1620

gatcctggag gccctgctga agcacatcag ctttgataag ttcgtgcaga tcagcctgaa 1680

ggccctgcgg aggatcgtgc cactgatgga gcagggcaag aggtacgacg aggcctgcgc 1740

cgaaatctac ggcgatcact atggcaagaa gaacacagag gagaaaatct acctgccccc 1800

tatccccgcc gatgagatca ggaaccctgt ggtgctgcgc gccctgtctc aggcaagaaa 1860

agtgatcaac ggagtggtgc gccggtacgg cagccccgcc agaatccaca tcgagacagc 1920

cagggaagtg ggcaagtcct ttaaggacag aaaggagatc gagaagaggc aggaggagaa 1980

cagaaaggat agggagaagg ccgccgccaa gttcagagag tactttccta atttcgtggg 2040

cgagccaaag tccaaggata tcctgaagct gaggctgtac gagcagcagc acggcaagtg 2100

tctgtattct ggcaaggaga tcaacctggg ccgcctgaat gagaagggct atgtggagat 2160

cgaccacgcc ctgccttttt ctcggacctg ggacgatagc ttcaacaata aggtgctggt 2220

gctgggctct gagaaccaga ataagggcaa ccagacaccc tacgagtatt tcaacggcaa 2280

ggacaatagc cgcgagtggc aggagtttaa ggcaagggtg gagacaagca ggttccctcg 2340

gtccaagaag cagagaatcc tgctgcagaa gtttgacgag gatggcttca aggagaggaa 2400

cctgaatgac acccgctacg tgaatcggtt tctgtgccag ttcgtggccg atagaatgag 2460

gctgaccggc aagggcaaga agagagtgtt tgcctccaac ggccagatca caaatctgct 2520

gaggggcttc tggggcctga gaaaggtgag ggcagagaac gacaggcacc acgcactgga 2580

tgcagtggtg gtggcatgtt ctaccgtggc catgcagcag aagatcacac gctttgtgcg 2640

gtataaggag atgaatgcct tcgacggcaa gaccatcgat aaggagacag gcgaggtgct 2700

gcaccagaag acacactttc ctcagccatg ggagttcttt gcccaggaag tgatgatccg 2760

ggtgtttggc aagcctgacg gcaagccaga gttcgaggag gccgataccc tggagaagct 2820

gagaacactg ctggcagaga agctgagctc caggcccgag gcagtgcacc agtacgtgac 2880

cccactgttc gtgtctagag cccccaacag gaagatgagc ggccagggcc acatggagac 2940

agtgaagtcc gccaagagac tggacgaggg cgtgtctgtg ctgagggtgc ctctgacaca 3000

gctgaagctg aaggatctgg agaagatggt gaatcgcgag cgggagccaa agctgtatga 3060

ggccctgaag gcaaggctgg aggcacacaa ggacgatcct gccaaggcct ttgccgagcc 3120

attctacaag tatgataagg ccggcaacag aacccagcag gtgaaggccg tgagggtgga 3180

gcaggtgcag aagacaggcg tgtgggtgcg caaccacaat ggcatcgccg acaatgctac 3240

catggtgcgg gtggacgtgt ttgagaaggg cgataagtac tatctggtgc ccatctacag 3300

ctggcaggtg gccaagggca tcctgcctga tagagccgtg gtgcagggca aggacgagga 3360

ggattggcag ctgatcgacg attccttcaa ctttaagttc tctctgcacc ccaatgacct 3420

ggtggaagtg atcaccaaga aggccaggat gtttggctac ttcgcctcct gccaccgcgg 3480

cacaggcaac atcaatatcc ggatccacga cctggatcac aagatcggca agaacggcat 3540

cctggagggc atcggcgtga agacagccct gagcttccag aagtatcaga tcgacgagct 3600

gggcaaggag atcagacctt gtaggctgaa gaagcgccca cccgtgcggg aggataagcg 3660

gcccgcagca accaagaagg caggacaggc caagaagaag aagtacccct atgacgtgcc 3720

tgactacgcc gggtatccct acgacgtgcc tgattacgcc gggtcctatc cctacgacgt 3780

gccagattac gctgcagctc cagcagcgaa gaaaagaagc tggattaaga tcttttccc 3840

tctgccaaaa attatgggga catcatgaag ccccttgagc atctgacttc tggctaataa 3900

aggaaattta ttttcattgc aatagtgtgt tggaattttt tgtgtctctc actcggcggc 3960

cgc 3963

<210> 231

<211> 835
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 231
tggcgtgatc tcccggcccc caggcgtcca gtacccacac cccagaaggc ttccaccttc 60

acgtggacgc gcaggctgcc ggtgggctcc cgttctctct ctctttctga ggctagagga 120

ctgagccagt ccttggctcc ccagagacat cacggcccgc agccccggag ccaagtgccc 180

cgagtcccag gcgtccatgt ccttcccgag gccgcgcgca cctctcctcg ccccgatggg 240

cacccactgc tctgcgtggc tgcggtggcc gctgttgccg ctgttgccgc cgctgctgct 300

gctgttgctg ctactgtgcc ccaccggcgc tggtgcccag gacgaggatg gagattatga 360

agagctgatg ctcgccctcc cgtcccagga ggatggcctg gctgatgagg ccgcacatgt 420

ggccaccgcc accttccgcc gttgctccaa ggtatgggtg ccaggcaacg gcgttgcttt 480

ggggttgggg tgatgctctt cgggggtctt ctctgctcat ctagccgtct ggtggtctct 540

aagtgcagcc ctgaggtgcg ggaggcgagg gcaagactta gtgctcagct gcaccttgtg 600

gcacagagtg atgggggagg ccacgtgcta aaggcactgc ggggcttggt tccaaaagtg 660

tgaggcgggg agcgggctac cagtgtggtc atgcagaaaa cgtgtcctcc gaagtaaagt 720

ggcatcggga ggctgagaac tctagtggca catctttctc aactggtcat ccagcagtca 780

tcctgggtgc ctgttggggt tctgtaggcc tcgatttagc ggaagggtgt caggg 835

<210> 232
<211> 863
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 232
ggaaagggaa aatgccaatg ctctgtctag gggttggata agccagtata ataaatgaaa 60

atggggctaa aatgagtgtt ctaaaatacc ttttgataag gctgcagaag gagcgggaga 120

aatggatatg aagtactggg ctctttaaaa atgattaaaa ttctgcttac atagtctaac 180

tcgcgacact gtaatttcat actgtagtaa ggatctcaag caggagagta taaaactcgg 240

gtgagcatgt ctttaatcta cctcgatgga aaatactccg aggcggatca caagcaataa 300

taacctgtag ttttgctgca taaaacccca gatgactacc tatcctccca ttttccttat 360

ttgcccctat taaaaaactt cccgacaaaa ccgaaaatct gtgggaagtc ttgtccctcc 420

aattttacac ctgttcaatt cccctgcagg acaacgccca cacaccaggt tagcctttaa 480

gcctgcccag aagactcccg cccatcttct agaaagactg gagttgcaga tcacgaggga 540

agaggggaa gggattctcc caggcccagg gcggtcctca gaagccagga ggcagcagag 600

aactcccaga aaggtattgc aacactcccc tccccccctcc ggagaagggt gcggccttct 660

ccccgcctac tccactgcag ctcccttact gataacaact cagagcgact ttgggagagc 720

aagtgcttcc tgcctccaaa acagcccaac tgagccctcg ctccttccct ccactccccg 780

gagtgcgcga tggaggtctg gctcagcacg cccctcttga ggcaactcaa gtcggaaacg 840

tgcttgcacc cgccccgcag ccg 863

<210> 233
<211> 3249
<212> DNA
<213> N. meningitidis

<400> 233
atggctgcct tcaaacctaa tccaatcaac tacatcctcg gcctcgatat cggcatcgca 60

tccgtcggct gggcgatggt agaaattgac gaagaagaaa accccatccg cctgattgat 120

ttgggcgtgc gcgtatttga gcgtgccgaa gtaccgaaaa caggcgactc ccttgccatg 180

gcaaggcgtt tggcgcgcag tgtccgccgc ctgacccgcc gtcgcgccca tcgcctgctt 240

cgggcccgcc gcctattgaa acgcgaaggc gtattacaag ccgctgattt tgacgaaaac 300

ggcttgatca aatccttacc gaatacacca tggcaacttc gcgcagccgc attagaccgc 360

aaactgacgc ctttagagtg gtcggcagtc ttgttgcatt taatcaaaca ccgcggttat 420

ttgtcgcaaa gaaaaaacga gggcgaaact gccgataaag agcttggcgc tttgcttaaa 480

ggcgtggcca acaatgccca tgccttacag acaggcgatt tccgcacacc ggccgaattg 540

gctttaaata aatttgagaa agaaagcggc catatccgca atcagcgcgg cgattattcg 600

catacgttca gccgcaaaga tttacaggcg gagctgattt tgctgtttga aaaacaaaaa 660

gaatttggca atccgcatgt ttcaggcggc cttaaagaag gtattgaaac cctactgatg 720

acgcaacgcc ctgccctgtc cggcgatgcc gttcaaaaaa tgttggggca ttgcaccttc 780

gaaccggcag agccgaaagc cgctaaaaac acctacacag ccgaacgttt catctggctg 840

accaagctga acaacctgcg tattttagag caaggcagcg agcggccatt gaccgatacc 900

gaacgcgcca cgcttatgga cgagccatac agaaaatcca aactgactta cgcacaagcc 960

cgtaagctgc tgggtttaga agataccgcc tttttcaaag gcttgcgcta tggtaaagac 1020

aatgccgaag cctcaacatt gatggaaatg aaggcctacc atgccatcag ccgtgcactg 1080

gaaaaagaag gattgaaaga taaaaaatcc ccattaaacc tttcttccga attacaagat 1140

gaaatcggca cggcattctc cctgttcaaa accgatgaag acattacagg ccgtctgaaa 1200

gaccgtgttc agcccgaaat cttagaagcg ctgttgaaac acatcagctt cgataagttc 1260

gtccaaattt ccttgaaagc attgcgccga attgtgcctc taatggagca aggcaaacgt 1320

tacgatgaag cctgcgccga aatctacgga gaccattacg gcaagaaaaa tacggaagaa 1380

aaaatttatc tgccgcccat ccctgccgac gagatccgca accccgtcgt cttgcgcgcc 1440

ttatctcaag cacgtaaggt cattaacggc gtggtacgcc gttacggctc cccagctcgt 1500

atccatattg aaacggcaag ggaagtaggt aaatcgttta aagaccgcaa agaaatcgaa 1560

aaacgccaag aagaaaaccg caaagaccgg gaaaaagccg ccgccaaatt ccgagagtat 1620

ttccccaatt ttgtcggcga acccaaatca aaagatattc tgaaactgcg cctgtacgag 1680

caacaacacg gcaaatgcct gtattcgggc aaagaaatca acttagtccg tctgaacgaa 1740

aaaggctatg tcgaaatcga ccatgccctg ccgttttcgc gcacatggga cgacagtttc 1800

aacaataaag tgctggtatt gggcagcgaa aaccaaaaca aaggcaatca aacccccttac 1860

gaatacttca acggcaaaga caacagccgc gaatggcagg aatttaaagc gcgtgtcgaa 1920

accagccgtt tcccgcgcag taaaaaacaa cggattctgc tgcaaaaatt cgatgaagac 1980

ggctttaaag aatgcaatct gaacgacacg cgctacgtca accgtttcct gtgccaattt 2040

gttgccgacc atatattgct gacaggtaaa gggaaaagac gtgtctttgc ctcaaacgga 2100

caaattacca atctgttgcg cggcttttgg ggattgcgca aagtgcgtgc ggaaaacgac 2160

cgccatcacg ccttggacgc tgtagtcgtt gcctgctcga ccgttgccat gcagcagaaa 2220

attacccgtt ttgtacgcta taaagagatg aacgcgtttg acggtaaaac catagacaaa 2280

gaaacaggaa aagtgctgca tcaaaaaaca cacttcccac aaccttggga atttttcgca 2340

caagaagtca tgattcgcgt cttcggcaaa ccggacggca aacccgaatt cgaagaagcc 2400

gataccccag aaaaactgcg cacgttgctt gccgaaaaat tatcatctcg ccccgaagcc 2460

gtacacgaat acgttacgcc actgtttgtt tcacgcgcgc ccaatcggaa gatgagcggt 2520

gcacataaag atactttgag atctgctaaa cgatttgtta aacataatga aaaaattagt 2580

gttaaacgag tatggttaac cgaaatcaag ttggccgacc ttgaaaatat ggttaattat 2640

aaaaatggta gagagattga attatatgag gctcttaagg cgcgtttaga ggcatatgga 2700

ggtaatgcta aacaagcatt tgaccctaag gacaatccgt tttataaaaa gggaggacaa 2760

ctggttaaag ctgtgagggt cgaaaaaacc caagagagcg gagtcttatt aaataaaaaa 2820

aatgcttata ccattgcaga taatggagac atggtacgtg ttgatgtgtt ctgtaaagta 2880

gataagaaag gaaaaaatca gtattttatt gttcccatct atgcttggca ggttgctgaa 2940

aacattcttc ccgatattga ttgtaaggga taccggattg atgatagcta tacattctgt 3000

tttagcttgc ataagtatga tctgattgct tttcaaaaag atgaaaaatc taaagtagaa 3060

ttcgcctact atatcaactg tgatagctct aatggacgat tctatttagc ttggcatgat 3120

aaaggctcta aagaacagca attccgtatt agcacccaaa atcttgtatt gatacaaaaa 3180

taccaagtta acgaactggg caaagaaatc agaccatgcc gtctgaaaaa acgcccacct 3240

gtccgttaa 3249

<210> 234
<211> 1082
<212> PRT
<213> N. meningitidis

<400> 234

Met Ala Ala Phe Lys Pro Asn Pro Ile Asn Tyr Ile Leu Gly Leu Asp
1 5 10 15

Ile Gly Ile Ala Ser Val Gly Trp Ala Met Val Glu Ile Asp Glu Glu
20 25 30

Glu Asn Pro Ile Arg Leu Ile Asp Leu Gly Val Arg Val Phe Glu Arg
35 40 45

Ala Glu Val Pro Lys Thr Gly Asp Ser Leu Ala Met Ala Arg Arg Leu
50 55 60

Ala Arg Ser Val Arg Arg Leu Thr Arg Arg Arg Ala His Arg Leu Leu
65 70 75 80

Arg Ala Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asp
85 90 95

Phe Asp Glu Asn Gly Leu Ile Lys Ser Leu Pro Asn Thr Pro Trp Gln
100 105 110

Leu Arg Ala Ala Ala Leu Asp Arg Lys Leu Thr Pro Leu Glu Trp Ser
115 120 125

Ala Val Leu Leu His Leu Ile Lys His Arg Gly Tyr Leu Ser Gln Arg
130 135 140

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

Gly Val Ala Asn Asn Ala His Ala Leu Gln Thr Gly Asp Phe Arg Thr
165 170 175

```
Pro Ala Glu Leu Ala Leu Asn Lys Phe Glu Lys Glu Ser Gly His Ile
            180                 185                 190

Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

Asp Arg Val Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
                405                 410                 415
```

```
Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
        595                 600                 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr
```

660 665 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr
675 680 685

Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
690 695 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705 710 715 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
725 730 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
740 745 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln
755 760 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
770 775 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785 790 795 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
805 810 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
820 825 830

Ala Pro Asn Arg Lys Met Ser Gly Ala His Lys Asp Thr Leu Arg Ser
835 840 845

Ala Lys Arg Phe Val Lys His Asn Glu Lys Ile Ser Val Lys Arg Val
850 855 860

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

Lys Asn Gly Arg Glu Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg Leu
885 890 895

Glu Ala Tyr Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn
900 905 910

Pro Phe Tyr Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val Glu
915 920 925

Lys Thr Gln Glu Ser Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr Thr
930 935 940

Ile Ala Asp Asn Gly Asp Met Val Arg Val Asp Val Phe Cys Lys Val
945 950 955 960

Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile Val Pro Ile Tyr Ala Trp
965 970 975

Gln Val Ala Glu Asn Ile Leu Pro Asp Ile Asp Cys Lys Gly Tyr Arg
980 985 990

Ile Asp Asp Ser Tyr Thr Phe Cys Phe Ser Leu His Lys Tyr Asp Leu
995 1000 1005

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

Tyr Ile Asn Cys Asp Ser Ser Asn Gly Arg Phe Tyr Leu Ala Trp
1025 1030 1035

His Asp Lys Gly Ser Lys Glu Gln Gln Phe Arg Ile Ser Thr Gln
1040 1045 1050

Asn Leu Val Leu Ile Gln Lys Tyr Gln Val Asn Glu Leu Gly Lys
1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1070 1075 1080

<210> 235
<211> 3423
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 235
atggccgcct tcaagcctaa cccaatcaat tacatcctgg gactggacat cggaatcgca 60

tccgtgggat gggctatggt ggagatcgac gaggaggaga atcctatccg gctgatcgat 120

ctgggcgtga gagtgtttga gagggccgag gtgccaaaga ccggcgattc tctggctatg 180

gcccggagac tggcacggag cgtgaggcgc ctgacacgga gaagggcaca caggctgctg 240

agggcacgcc ggctgctgaa gagagagggc gtgctgcagg cagcagactt cgatgagaat 300

ggcctgatca agagcctgcc aaacaccccc tggcagctga gagcagccgc cctggacagg 360

aagctgacac cactggagtg gtctgccgtg ctgctgcacc tgatcaagca ccgcggctac 420

ctgagccagc ggaagaacga gggagagaca gcagacaagg agctgggcgc cctgctgaag 480

ggagtggcca acaatgccca cgccctgcag accggcgatt tcaggacacc tgccgagctg 540

gccctgaata agtttgagaa ggagtccggc cacatcagaa accagagggg cgactatagc 600

cacaccttct cccgcaagga tctgcaggcc gagctgatcc tgctgttcga gaagcagaag 660

gagtttggca atccacacgt gagcggaggc ctgaaggagg gaatcgagac cctgctgatg 720

acacagaggc ctgccctgtc cggcgacgca gtgcagaaga tgctgggaca ctgcaccttc 780

gagcctgcag agccaaaggc cgccaagaac acctacacag ccgagcggtt tatctggctg 840

acaaagctga acaatctgag aatcctggag cagggatccg agaggccact gaccgacaca 900

gagagggcca ccctgatgga tgagccttac cggaagtcta agctgacata tgcccaggcc 960

agaaagctgc tgggcctgga ggacaccgcc ttctttaagg gcctgagata cggcaaggat 1020

aatgccgagg cctccacact gatggagatg aaggcctatc acgccatctc tcgcgccctg 1080

gagaaggagg gcctgaagga caagaagtcc cccctgaacc tgagctccga gctgcaggat 1140

gagatcggca ccgccttctc tctgtttaag accgacgagg atatcacagg ccgcctgaag 1200

gacagggtgc agcctgagat cctggaggcc ctgctgaagc acatctcttt cgataagttt 1260

gtgcagatca gcctgaaggc cctgagaagg atcgtgccac tgatggagca gggcaagcgg 1320

tacgacgagg cctgcgccga gatctacggc gatcactatg gcaagaagaa cacagaggag 1380

aagatctatc tgcccccctat ccctgccgac gagatcagaa atcctgtggt gctgagggcc 1440

ctgtcccagg caagaaaagt gatcaacgga gtggtgcgcc ggtacggatc tccagcccgg 1500

atccacatcg agaccgccag agaagtgggc aagagcttca aggaccggaa ggagatcgag 1560

aagagacagg aggagaatcg caaggatcgg gagaaggccg ccgccaagtt tagggagtac 1620

ttccctaact ttgtgggcga gccaaagtct aaggacatcc tgaagctgcg cctgtacgag 1680

cagcagcacg gcaagtgtct gtatagcggc aaggagatca atctggtgcg gctgaacgag 1740

aagggctatg tggagatcga tcacgccctg cctttctcca gaacctggga cgattctttt 1800

aacaataagg tgctggtgct gggcagcgag aaccagaata agggcaatca gacaccatac 1860

gagtatttca atggcaagga caactccagg gagtggcagg agttcaaggc ccgcgtggag 1920

acctctagat ttcccaggag caagaagcag cggatcctgc tgcagaagtt cgacgaggat 1980

ggctttaagg agtgcaacct gaatgacacc agatacgtga accggttcct gtgccagttt 2040

gtggccgatc acatcctgct gaccggcaag ggcaagagaa gggtgttcgc ctctaatggc 2100

cagatcacaa acctgctgag gggattttgg ggactgagga aggtgcgggc agagaatgac 2160

agacaccacg cactggatgc agtggtggtg gcatgcagca ccgtggcaat gcagcagaag 2220

atcacaagat tcgtgaggta taaggagatg aacgcctttg acggcaagac catcgataag 2280

gagacaggca aggtgctgca ccagaagacc cacttccccc agccttggga gttctttgcc 2340

caggaagtga tgatccgggt gttcggcaag ccagacggca agcctgagtt tgaggaggcc 2400

gataccccag agaagctgag gacactgctg gcagagaagc tgtctagcag gccagaggca 2460

gtgcacgagt acgtgacccc actgttcgtg tccagggcac ccaatcggaa gatgtctggc 2520

gcccacaagg acacactgag aagcgccaag aggtttgtga agcacaacga gaagatctcc 2580

gtgaagagag tgtggctgac cgagatcaag ctggccgatc tggagaacat ggtgaattac 2640

aagaacggca gggagatcga gctgtatgag gccctgaagg caaggctgga ggcctacgga 2700

ggaaatgcca agcaggcctt cgacccaaag gataacccct tttataagaa gggaggacag 2760

ctggtgaagg ccgtgcgggt ggagaagacc caggagagcg gcgtgctgct gaataagaag 2820

aacgcctaca caatcgccga caatggcgat atggtgagag tggacgtgtt ctgtaaggtg 2880

gataagaagg gcaagaatca gtactttatc gtgcctatct atgcctggca ggtggccgag 2940

aacatcctgc cagacatcga ttgcaagggc tacagaatcg acgatagcta tacattctgt 3000

ttttccctgc acaagtatga cctgatcgcc ttccagaagg atgagaagtc caaggtggag 3060

tttgcctact atatcaattg cgactcctct aacggcaggt tctacctggc ctggcacgat 3120

aagggcagca aggagcagca gtttcgcatc tccacccaga atctggtgct gatccagaag 3180

tatcaggtga acgagctggg caaggagatc aggccatgtc ggctgaagaa gcgcccaccc 3240

gtgcggggca ccggcgggcc caagaagaag aggaaggtat acccatacga tgttcctgac 3300

tatgcgggct atccctatga cgtcccggac tatgcaggat cgtatcctta tgacgttcca 3360

gattacgctg gatccgccgc tccggcagct aagaaaaaga aactggattt cgaatccgga 3420

taa 3423

<210> 236
<211> 1141
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 236

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 Ile Ile
            180                 185                 190

Leu Arg Asn Gln Arg Gly Asp Tyr Ser His Thr Phe Ser Arg Lys Asp
        195                 200                 205

Leu Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly
    210                 215                 220

Asn Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu
225                 230                 235                 240

Met Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu
                245                 250                 255
```

```
Gly His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr
           260                 265                 270

Tyr Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg
       275                 280                 285

Ile Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala
   290                 295                 300

Thr Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln
305                 310                 315                 320

Ala Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu
               325                 330                 335

Arg Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys
           340                 345                 350

Ala Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp
       355                 360                 365

Lys Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly
   370                 375                 380

Thr Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu
385                 390                 395                 400

Lys Asp Arg Val Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile
               405                 410                 415

Ser Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile
           420                 425                 430

Val Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu
       435                 440                 445

Ile Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr
   450                 455                 460

Leu Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg
465                 470                 475                 480

Ala Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr
               485                 490                 495
```

```
Gly Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys
            500                 505                 510

Ser Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg
        515                 520                 525

Lys Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn
    530                 535                 540

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

Glu Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu
                565                 570                 575

Val Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro
            580                 585                 590

Phe Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu
        595                 600                 605

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

Asn Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val
625                 630                 635                 640

Glu Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln
                645                 650                 655

Lys Phe Asp Glu Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg
            660                 665                 670

Tyr Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu
        675                 680                 685

Thr Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr
    690                 695                 700

Asn Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn
705                 710                 715                 720

Asp Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val
                725                 730                 735
```

```
Ala Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn
            740                 745                 750

Ala Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His
        755                 760                 765

Gln Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val
    770                 775                 780

Met Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu
785                 790                 795                 800

Ala Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser
                805                 810                 815

Ser Arg Pro Glu Ala Val His Glu Tyr Val Phe Pro Leu Phe Val Ser
            820                 825                 830

Arg Ala Pro Asn Arg Lys Met Ser Gly Ala His Lys Asp Thr Leu Arg
        835                 840                 845

Ser Ala Lys Arg Phe Val Lys His Asn Glu Lys Ile Ser Val Lys Arg
    850                 855                 860

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

Tyr Lys Asn Gly Arg Glu Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg
                885                 890                 895

Leu Glu Ala Tyr Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp
            900                 905                 910

Asn Pro Phe Tyr Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val
        915                 920                 925

Glu Lys Thr Gln Glu Ser Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr
    930                 935                 940

Thr Ile Ala Asp Asn Gly Asp Met Val Arg Val Asp Val Phe Cys Lys
945                 950                 955                 960

Val Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile Val Pro Ile Tyr Ala
                965                 970                 975

Trp Gln Val Ala Glu Asn Ile Leu Pro Asp Ile Asp Cys Lys Gly Tyr
```

980 985 990

Arg Ile Asp Asp Ser Tyr Thr Phe Cys Phe Ser Leu His Lys Tyr Asp
995 1000 1005

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

Tyr Tyr Ile Asn Cys Asp Ser Ser Asn Gly Arg Phe Tyr Leu Ala
1025 1030 1035

Trp His Asp Lys Gly Ser Lys Glu Gln Gln Phe Arg Ile Ser Thr
1040 1045 1050

Gln Asn Leu Val Leu Ile Gln Lys Tyr Gln Val Asn Glu Leu Gly
1055 1060 1065

Lys Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1070 1075 1080

Gly Thr Gly Gly Pro Lys Lys Lys Arg Lys Val Tyr Pro Tyr Asp
1085 1090 1095

Val Pro Asp Tyr Ala Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1100 1105 1110

Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Ser Ala Ala
1115 1120 1125

Pro Ala Ala Lys Lys Lys Lys Leu Asp Phe Glu Ser Gly
1130 1135 1140

<210> 237
<211> 1054
<212> PRT
<213> H. parainfluenzae

<400> 237

Met Glu Asn Lys Asn Leu Asn Tyr Ile Leu Gly Leu Asp Leu Gly Ile
1 5 10 15

Ala Ser Val Gly Trp Ala Val Val Glu Ile Asp Glu Lys Glu Asn Pro
20 25 30

Leu Arg Leu Ile Asp Val Gly Val Arg Thr Phe Glu Arg Ala Glu Val
35 40 45

```
Pro Lys Thr Gly Glu Ser Leu Ala Leu Ser Arg Arg Leu Ala Arg Ser
    50                  55                  60

Ala Arg Arg Leu Thr Gln Arg Arg Val Ala Arg Leu Lys Lys Ala Lys
65                  70                  75                  80

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

Pro His Gln Val Trp Gln Leu Arg Val Glu Gly Leu Asp His Lys Leu
            100                 105                 110

Glu Arg Gln Glu Trp Ala Ala Val Leu Leu His Leu Ile Lys His Arg
        115                 120                 125

Gly Tyr Leu Ser Gln Arg Lys Asn Glu Ser Lys Ser Glu Asn Lys Glu
    130                 135                 140

Leu Gly Ala Leu Leu Ser Gly Val Asp Asn Asn His Lys Leu Leu Gln
145                 150                 155                 160

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

Val Glu Glu Gly His Ile Arg Asn Gln Gln Gly Ala Tyr Thr His Thr
            180                 185                 190

Phe Ser Arg Leu Asp Leu Leu Ala Glu Met Glu Leu Leu Phe Ser Arg
        195                 200                 205

Gln Gln His Phe Gly Asn Pro Phe Ala Ser Glu Lys Leu Leu Glu Asn
    210                 215                 220

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

Ile Leu Lys Met Leu Gly Lys Cys Thr Phe Glu Asp Glu Tyr Lys Ala
                245                 250                 255

Ala Lys Asn Thr Tyr Ser Ala Glu Arg Phe Val Trp Ile Thr Lys Leu
            260                 265                 270

Asn Asn Leu Arg Ile Gln Glu Asn Gly Leu Glu Arg Ala Leu Asn Asp
        275                 280                 285
```

```
Asn Glu Arg Leu Ala Leu Met Glu Gln Pro Tyr Asp Lys Asn Arg Leu
    290                 295                 300

Phe Tyr Ser Gln Val Arg Ser Ile Leu Lys Leu Ser Asp Glu Ala Ile
305                 310                 315                 320

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

Lys Ala Val Leu Met Glu Met Lys Ala Tyr His Gln Ile Arg Lys Val
            340                 345                 350

Leu Glu Gly Asn Asn Leu Lys Ala Glu Trp Ala Glu Leu Lys Ala Asn
        355                 360                 365

Pro Thr Leu Leu Asp Glu Ile Gly Thr Ala Phe Ser Leu Tyr Lys Thr
    370                 375                 380

Asp Glu Asp Ile Ser Ala Tyr Leu Ala Gly Lys Leu Ser Gln Pro Val
385                 390                 395                 400

Leu Asn Ala Leu Leu Glu Asn Leu Ser Phe Asp Lys Phe Ile Gln Leu
                405                 410                 415

Ser Leu Lys Ala Leu Tyr Lys Leu Leu Pro Leu Met Gln Gln Gly Leu
            420                 425                 430

Arg Tyr Asp Glu Ala Cys Arg Glu Ile Tyr Gly Asp His Tyr Gly Lys
        435                 440                 445

Lys Thr Glu Glu Asn His His Phe Leu Pro Gln Ile Pro Ala Asp Glu
    450                 455                 460

Ile Arg Asn Pro Val Val Leu Arg Thr Leu Thr Gln Ala Arg Lys Val
465                 470                 475                 480

Ile Asn Gly Val Val Arg Leu Tyr Gly Ser Pro Ala Arg Ile His Ile
                485                 490                 495

Glu Thr Gly Arg Glu Val Gly Lys Ser Tyr Lys Asp Arg Arg Glu Leu
            500                 505                 510

Glu Lys Arg Gln Glu Glu Asn Arg Lys Gln Arg Glu Asn Ala Ile Lys
        515                 520                 525

Glu Phe Lys Glu Tyr Phe Pro His Phe Ala Gly Glu Pro Lys Ala Lys
```

530 535 540

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

Tyr Ser Gly Lys Pro Ile Glu Leu His Arg Leu Leu Glu Lys Gly Tyr
565 570 575

Val Glu Val Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp Ser
580 585 590

Phe Asn Asn Lys Val Leu Val Leu Ala Asn Glu Asn Gln Asn Lys Gly
595 600 605

Asn Leu Thr Pro Phe Glu Trp Leu Asp Gly Lys His Asn Ser Glu Arg
610 615 620

Trp Arg Ala Phe Lys Ala Leu Val Glu Thr Ser Ala Phe Pro Tyr Ala
625 630 635 640

Lys Lys Gln Arg Ile Leu Ser Gln Lys Leu Asp Glu Lys Gly Phe Ile
645 650 655

Glu Arg Asn Leu Asn Asp Thr Arg Tyr Val Ala Arg Phe Leu Cys Asn
660 665 670

Phe Ile Ala Asp Asn Met His Leu Thr Gly Glu Gly Lys Arg Lys Val
675 680 685

Phe Ala Ser Asn Gly Gln Ile Thr Ala Leu Leu Arg Ser Arg Trp Gly
690 695 700

Leu Ala Lys Ser Arg Glu Asp Asn Asp Arg His His Ala Leu Asp Ala
705 710 715 720

Val Val Val Ala Cys Ser Thr Val Ala Met Gln Gln Lys Ile Thr Arg
725 730 735

Phe Val Arg Phe Glu Ala Gly Asp Val Phe Thr Gly Glu Arg Ile Asp
740 745 750

Arg Glu Thr Gly Glu Ile Ile Pro Leu His Phe Pro Thr Pro Trp Gln
755 760 765

Phe Phe Lys Gln Glu Val Glu Ile Arg Ile Phe Ser Asp Asn Pro Lys
770 775 780

```
Leu Glu Leu Glu Asn Arg Leu Pro Asp Arg Pro Gln Ala Asn His Glu
785                 790                 795                 800

Phe Val Gln Pro Leu Phe Val Ser Arg Met Pro Thr Arg Lys Met Thr
                805                 810                 815

Gly Gln Gly His Met Glu Thr Val Lys Ser Ala Lys Arg Leu Asn Glu
            820                 825                 830

Gly Ile Ser Val Ile Lys Met Pro Leu Thr Lys Leu Lys Leu Lys Asp
        835                 840                 845

Leu Glu Leu Met Val Asn Arg Glu Arg Glu Lys Asp Leu Tyr Asp Thr
    850                 855                 860

Leu Lys Ala Arg Leu Glu Ala Phe Asn Asp Asp Pro Ala Lys Ala Phe
865                 870                 875                 880

Ala Glu Pro Phe Ile Lys Lys Gly Gly Ala Ile Val Lys Ser Val Arg
                885                 890                 895

Val Glu Gln Ile Gln Lys Ser Gly Val Leu Val Arg Glu Gly Asn Gly
            900                 905                 910

Val Ala Asp Asn Ala Ser Met Val Arg Val Asp Val Phe Thr Lys Gly
        915                 920                 925

Gly Lys Tyr Phe Leu Val Pro Ile Tyr Thr Trp Gln Val Ala Lys Gly
    930                 935                 940

Ile Leu Pro Asn Lys Ala Ala Thr Gln Tyr Lys Asp Glu Glu Asp Trp
945                 950                 955                 960

Glu Val Met Asp Asn Ser Ala Thr Phe Lys Phe Ser Leu His Pro Asn
                965                 970                 975

Asp Leu Val Lys Leu Val Thr Lys Lys Lys Thr Ile Leu Gly Tyr Phe
            980                 985                 990

Asn Gly Leu Asn Arg Ala Thr Gly  Asn Ile Asp Ile Lys  Glu His Asp
        995                 1000                1005

Leu Asp  Lys Ser Lys Gly Lys  Gln Gly Ile Phe Glu  Gly Val Gly
    1010                1015                1020
```

Ile Lys Leu Ala Leu Ser Phe Glu Lys Tyr Gln Val Asp Glu Leu
1025 1030 1035

Gly Lys Asn Ile Arg Leu Cys Lys Pro Ser Lys Arg Gln Pro Val
1040 1045 1050

Arg

<210> 238
<211> 3192
<212> DNA
<213> S. muelleri

<400> 238
atggaaaaat ttcactatgt attgggtttg gatttgggta tcgcctctgt ggggtgggct 60

gccattgaaa ttgacaagga aaccgaaaca tcaatcggtt tattggattg cggtgtcaga 120

acatttgaac gtgcagaagt acccaaaaca ggcgattctc ttgccaaagc tcgccgtgaa 180

gccagaagta ctcgccgttt aattcgcaga cgttcgcatc gcttattacg tttaaaacgt 240

ttattgaaac gtgaaatttt caggcagcct gaaacgttta aagacttacc aatcaatgct 300

tggcaattgc gtgttaaagg cttggatagt cggttgaatg aatatgaatg ggcggccgtt 360

ttattgcatt tggtgaagca tcgcggttat ttatcgcaac gcaaaagcga aatgagcgaa 420

acagacagca aatctgaaat gggcagatta ctggcaggtg tggcggaaaa tcaccaactt 480

ttacaacaag aacaatatcg tacaccagcc gaattagcac tcaaaaaatt tgtgaaacat 540

tttcgcaata aaggtggcga ttatgcacac actttcaacc gtttggattt gcaagccgaa 600

ttgcatttat tgtttcaaaa acaacgtgaa ttaggcaatc cattcacttc accagaattg 660

gaacggcaag ttgatgattt gttgatgacg cagcgcagtg ctttacaagg tgatgcgatt 720

ttgaaaatgt tgggtcattg tgggtttgaa cctgaacaat tcaaagcagc gaaaaacaca 780

ttcagtgccg aacgttttat ttggttgaca aaactcaata atcttcgcat tcaagaccaa 840

ggcaaagaac gtgcgttaac tgccgatgag cgtaccaaat tgttggacga gccttataaa 900

aaaagtaaat tgacttacgc acaagttcgc aaattattaa gcttgcctca aactgctatt 960

tttaaaggtt tgcgttatga tttggaacat gacaaaaaag cagaaaacag tacgttgatg 1020

gaaatgaaat cctatcacaa catccgccaa acattggaaa aatcaggttt gaaaacagaa 1080

tggcaaagta ttgccacgca gcctgaaatt ttagatgcaa ttggcacggc gttttccatt 1140

tataaaaccg atgaagatat ttcgcatgaa ttaaaaacgt gcaggctgcc tgaaaacgta 1200

ttgaatgaat tactgaaaaa catcaatttt gatggattca ttcaattatc gttgactgca 1260

ttacgcaaaa ttttgccctt gatggaacaa ggctaccgtt atgatgaagc gtgtacccaa 1320

atttacggta atcatcattc aggcagcttg caacaagaat caaagcaatt tttgccacat 1380

attccgattg atgatgtccg aaatcctgtg gtgttccgta ctttgaccca agcaagaaaa 1440

gtggtgaatg cgattattcg tcggtatggt tcgccagctc gtgtgcatat tgagatggcg 1500

cgtgaattgg gtaagtctaa atcagaccgt gaccgaattg aaaaacaaca acaaaaaaat 1560

aaaaaagaac gtgaaaacgc agtcgccaaa ttcaaagaag atttccctga ttttgtgggc 1620

gaacccagag ggaaagatat tttgaaaatg cgtttgtatg aacaacaaca cggcaaatgt 1680

ttgtattcgg gtcatgatat tgatattaat cgattgaatg aaaaaggtta tgttgaaatt 1740

gaccatgccc tgccattttc acggacttgg gatgatagtc aaaataataa agttttggta 1800

cttggcagcg aaaaccaaaa taaacgcaat caaacgcctg atgaatattt ggacggtgca 1860

aacaatagcc aacgttggct tgaatttcaa gcgcgtgtac aaacttgtca tttttcttac 1920

ggtaaaaaac aacgcattca attagccaaa ttagacgatg aaaccgaaaa aggattttta 1980

gaacgcaatc taaatgatac gcggtatatt gctcgtttta tgtgtcaatt tgtccaagaa 2040

aatttatatt tgacaggtaa aggaaaacgt cttgtttttg catcaaacgg cggaatgacc 2100

gcaacattga gaaatttatg gggtttgaga aaagtccgtg aagacaacga ccgccatcat 2160

gctcttgatg cgattgtggt ggcgtgttcc actgcttcta tgcaacaaaa aataaccaaa 2220

gcatttcaac gccatgaaag cattgaatat gtggataccg aaacgggcga agtaaaattt 2280

cgtattccac agccttggga ttttttccgt caggaagtga tgattcgtgt gttcagcgac 2340

caaccgtgtg aagatttggt agaaaaattg tcggctcgtc ccgaagcttt gcatgacaac 2400

gtaacgcctt tatttgtctc gcgtgcacca aatcgcaaaa tgtcggggca agggcatttg 2460

gaaaccatca aatctgcaaa aaggctgtct gaagaaaaca gtatggttaa aaaaccatta 2520

accacattga aattaaaaga tattccagaa atcgtaggct acccgagtcg tgaacctcaa 2580

ttgtatgccg cattgaaaac acgtttagaa acgcatgatg atgacccaat taaagccttt 2640

gccaaaccct tttacaaacc caataaaaat ggtgaattgg gggcgttggt tcgatcggtg 2700

cgtgtgaaag gtgtacaaaa tacgggtgta atggttcatg atggcaaagg cattgccgat 2760

aatgccacaa tggttcgtgt tgatgtctat accaaagcgg gcaaaaatta ccttgttcct 2820

gtgtatgttt ggcaggtggc tcaaggaatt ttgccaaatc gggcggttac ttctggcaaa 2880

agtgaagcag attgggattt aattgatgaa agttttgaat ttaaattttc gctgtctcgt 2940

ggggatttag tggaaatgat tagcaataaa ggaagaattt ttggttatta caatgggtta 3000

gatcgtgcaa atggaagtat tgggattcgt gaacatgatt tggaaaagtc caaaggaaaa 3060

gatggtgttc atcgtgttgg cgtgaaaacc gccaccgcat tcaacaaata ccacgttgac 3120

ccacttggta aagaaattca tcggtgttca tctgaaccac gccccacatt aaaaatcaaa 3180

tccaagaaat aa 3192

<210> 239
<211> 1063
<212> PRT
<213> S. muelleri

<400> 239

Met Glu Lys Phe His Tyr Val Leu Gly Leu Asp Leu Gly Ile Ala Ser
1 5 10 15

Val Gly Trp Ala Ala Ile Glu Ile Asp Lys Glu Thr Glu Thr Ser Ile
20 25 30

Gly Leu Leu Asp Cys Gly Val Arg Thr Phe Glu Arg Ala Glu Val Pro
35 40 45

Lys Thr Gly Asp Ser Leu Ala Lys Ala Arg Arg Glu Ala Arg Ser Thr
50 55 60

Arg Arg Leu Ile Arg Arg Arg Ser His Arg Leu Leu Arg Leu Lys Arg
65 70 75 80

Leu Leu Lys Arg Glu Ile Phe Arg Gln Pro Glu Thr Phe Lys Asp Leu
85 90 95

Pro Ile Asn Ala Trp Gln Leu Arg Val Lys Gly Leu Asp Ser Arg Leu
100 105 110

Asn Glu Tyr Glu Trp Ala Ala Val Leu Leu His Leu Val Lys His Arg
115 120 125

Gly Tyr Leu Ser Gln Arg Lys Ser Glu Met Ser Glu Thr Asp Ser Lys
130 135 140

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

Leu Gln Gln Glu Gln Tyr Arg Thr Pro Ala Glu Leu Ala Leu Lys Lys
165 170 175

Phe Val Lys His Phe Arg Asn Lys Gly Gly Asp Tyr Ala His Thr Phe
180 185 190

```
Asn Arg Leu Asp Leu Gln Ala Glu Leu His Leu Leu Phe Gln Lys Gln
        195                 200                 205

Arg Glu Leu Gly Asn Pro Phe Thr Ser Pro Glu Leu Glu Arg Gln Val
    210                 215                 220

Asp Asp Leu Leu Met Thr Gln Arg Ser Ala Leu Gln Gly Asp Ala Ile
225                 230                 235                 240

Leu Lys Met Leu Gly His Cys Gly Phe Glu Pro Glu Gln Phe Lys Ala
                245                 250                 255

Ala Lys Asn Thr Phe Ser Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu
            260                 265                 270

Asn Asn Leu Arg Ile Gln Asp Gln Gly Lys Glu Arg Ala Leu Thr Ala
        275                 280                 285

Asp Glu Arg Thr Lys Leu Leu Asp Glu Pro Tyr Lys Lys Ser Lys Leu
    290                 295                 300

Thr Tyr Ala Gln Val Arg Lys Leu Leu Ser Leu Pro Gln Thr Ala Ile
305                 310                 315                 320

Phe Lys Gly Leu Arg Tyr Asp Leu Glu His Asp Lys Lys Ala Glu Asn
                325                 330                 335

Ser Thr Leu Met Glu Met Lys Ser Tyr His Asn Ile Arg Gln Thr Leu
            340                 345                 350

Glu Lys Ser Gly Leu Lys Thr Glu Trp Gln Ser Ile Ala Thr Gln Pro
        355                 360                 365

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

Glu Asp Ile Ser His Glu Leu Lys Thr Cys Arg Leu Pro Glu Asn Val
385                 390                 395                 400

Leu Asn Glu Leu Leu Lys Asn Ile Asn Phe Asp Gly Phe Ile Gln Leu
                405                 410                 415

Ser Leu Thr Ala Leu Arg Lys Ile Leu Pro Leu Met Glu Gln Gly Tyr
            420                 425                 430

Arg Tyr Asp Glu Ala Cys Thr Gln Ile Tyr Gly Asn His His Ser Gly
```

435 440 445

Ser Leu Gln Gln Glu Ser Lys Gln Phe Leu Pro His Ile Pro Ile Asp
450 455 460

Asp Val Arg Asn Pro Val Val Phe Arg Thr Leu Thr Gln Ala Arg Lys
465 470 475 480

Val Val Asn Ala Ile Ile Arg Arg Tyr Gly Ser Pro Ala Arg Val His
485 490 495

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

Ile Glu Lys Gln Gln Gln Lys Asn Lys Lys Glu Arg Glu Asn Ala Val
515 520 525

Ala Lys Phe Lys Glu Asp Phe Pro Asp Phe Val Gly Glu Pro Arg Gly
530 535 540

Lys Asp Ile Leu Lys Met Arg Leu Tyr Glu Gln Gln His Gly Lys Cys
545 550 555 560

Leu Tyr Ser Gly His Asp Ile Asp Ile Asn Arg Leu Asn Glu Lys Gly
565 570 575

Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp
580 585 590

Ser Gln Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys
595 600 605

Arg Asn Gln Thr Pro Asp Glu Tyr Leu Asp Gly Ala Asn Asn Ser Gln
610 615 620

Arg Trp Leu Glu Phe Gln Ala Arg Val Gln Thr Cys His Phe Ser Tyr
625 630 635 640

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

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

Phe Met Cys Gln Phe Val Gln Glu Asn Leu Tyr Leu Thr Gly Lys Gly
675 680 685

```
Lys Arg Leu Val Phe Ala Ser Asn Gly Gly Met Thr Ala Thr Leu Arg
    690                 695                 700

Asn Leu Trp Gly Leu Arg Lys Val Arg Glu Asp Asn Asp Arg His His
705                 710                 715                 720

Ala Leu Asp Ala Ile Val Val Ala Cys Ser Thr Ala Ser Met Gln Gln
                725                 730                 735

Lys Ile Thr Lys Ala Phe Gln Arg His Glu Ser Ile Glu Tyr Val Asp
            740                 745                 750

Thr Glu Thr Gly Glu Val Lys Phe Arg Ile Pro Gln Pro Trp Asp Phe
        755                 760                 765

Phe Arg Gln Glu Val Met Ile Arg Val Phe Ser Asp Gln Pro Cys Glu
    770                 775                 780

Asp Leu Val Glu Lys Leu Ser Ala Arg Pro Glu Ala Leu His Asp Asn
785                 790                 795                 800

Val Thr Pro Leu Phe Val Ser Arg Ala Pro Asn Arg Lys Met Ser Gly
                805                 810                 815

Gln Gly His Leu Glu Thr Ile Lys Ser Ala Lys Arg Leu Ser Glu Glu
            820                 825                 830

Asn Ser Met Val Lys Lys Pro Leu Thr Thr Leu Lys Leu Lys Asp Ile
        835                 840                 845

Pro Glu Ile Val Gly Tyr Pro Ser Arg Glu Pro Gln Leu Tyr Ala Ala
    850                 855                 860

Leu Lys Thr Arg Leu Glu Thr His Asp Asp Asp Pro Ile Lys Ala Phe
865                 870                 875                 880

Ala Lys Pro Phe Tyr Lys Pro Asn Lys Asn Gly Glu Leu Gly Ala Leu
                885                 890                 895

Val Arg Ser Val Arg Val Lys Gly Val Gln Asn Thr Gly Val Met Val
            900                 905                 910

His Asp Gly Lys Gly Ile Ala Asp Asn Ala Thr Met Val Arg Val Asp
        915                 920                 925
```

Val Tyr Thr Lys Ala Gly Lys Asn Tyr Leu Val Pro Val Tyr Val Trp
930 935 940

Gln Val Ala Gln Gly Ile Leu Pro Asn Arg Ala Val Thr Ser Gly Lys
945 950 955 960

Ser Glu Ala Asp Trp Asp Leu Ile Asp Glu Ser Phe Glu Phe Lys Phe
965 970 975

Ser Leu Ser Arg Gly Asp Leu Val Glu Met Ile Ser Asn Lys Gly Arg
980 985 990

Ile Phe Gly Tyr Tyr Asn Gly Leu Asp Arg Ala Asn Gly Ser Ile Gly
995 1000 1005

Ile Arg Glu His Asp Leu Glu Lys Ser Lys Gly Lys Asp Gly Val
1010 1015 1020

His Arg Val Gly Val Lys Thr Ala Thr Ala Phe Asn Lys Tyr His
1025 1030 1035

Val Asp Pro Leu Gly Lys Glu Ile His Arg Cys Ser Ser Glu Pro
1040 1045 1050

Arg Pro Thr Leu Lys Ile Lys Ser Lys Lys
1055 1060

<210> 240
<211> 3293
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 240
atggagaagt tccactacgt gctgggactg gatctgggaa tcgcaagcgt gggatgggca 60

gcaatcgaga tcgataagga gaccgagaca tccatcggcc tgctggactg cggcgtgagg 120

acctttgaga gggcagaggt gcctaagaca ggcgacagcc tggcaaaggc aaggagagag 180

gcaaggtcta caaggcgcct gatccggaga aggagccaca ggctgctgcg gctgaagaga 240

ctgctgaagc gggagatctt tagacagcca gagaccttca aggatctgcc catcaacgca 300

tggcagctga gggtgaaggg actggactct cggctgaatg agtacgagtg ggcagccgtg 360

ctgctgcacc tggtgaagca caggggctat ctgagccagc gcaagtccga gatgtctgag 420

accgactcta agagcgagat gggcaggctg ctggcaggag tggccgagaa ccaccagctg 480

ctgcagcagg agcagtacag gaccccagca gagctggccc tgaagaagtt tgtgaagcac 540

ttccgcaaca agggcggcga ttatgcccac acattcaata ggctggacct gcaggcagag 600

ctgcacctgc tgtttcagaa gcagagagag ctgggcaacc ccttcacctc tcctgagctg 660

gagcgccagg tggacgatct gctgatgaca cagcggagcg ccctgcaggg cgatgcaatc 720

ctgaagatgc tgggccactg tggctttgag cctgagcagt tcaaggccgc caagaatacc 780

tttagcgccg agagattcat ctggctgaca aagctgaaca atctgaggat ccaggaccag 840

ggcaaggaga gagccctgac cgccgatgag aggacaaagc tgctggacga gccttacaag 900

aagtctaagc tgacctatgc ccaggtgagg aagctgctga gcctgcctca gacagccatc 960

ttcaagggcc tgcgctacga tctggagcac gacaagaagg ccgagaactc taccctgatg 1020

gagatgaaga gctatcacaa tatccggcag acactggaga agtccggcct gaagaccgag 1080

tggcagtcta tcgccacaca gccagagatc ctggacgcaa tcggaaccgc cttttccatc 1140

tacaagacag atgaggacat ctctcacgag ctgaagacct gcagactgcc tgagaacgtg 1200

ctgaatgagc tgctgaagaa catcaatttt gatggcttca tccagctgag cctgaccgcc 1260

ctgcgcaaga tcctgccact gatggagcag ggctaccggt atgacgaggc ctgtacacag 1320

atctacggca accaccactc cggctctctg cagcaggagt ccaagcagtt tctgcctcac 1380

atcccaatcg acgatgtgcg gaacccagtg gtgttcagaa ccctgacaca ggccaggaag 1440

gtggtgaatg ccatcatccg ccggtatgga tctccagcaa gggtgcacat cgagatggca 1500

agggagctgg gcaagagcaa gtccgataga gacaggatcg agaagcagca gcagaagaac 1560

aagaaggaga gggagaatgc cgtggccaag ttcaaggagg attttccaga cttcgtggga 1620

gagcctaggg gcaaggatat cctgaagatg cggctgtacg agcagcagca cggcaagtgc 1680

ctgtattccg gccacgatat cgacatcaac cggctgaatg agaagggcta cgtggagatc 1740

gaccacgccc tgccttttag cagaacctgg gacgattccc agaacaataa ggtgctggtg 1800

ctgggcagcg agaaccagaa taagcgcaat cagacaccag atgagtacct ggacggcgcc 1860

aacaattccc agagatggct ggagtttcag gccagggtgc agacctgcca cttctcttat 1920

ggcaagaagc agaggatcca gctggccaag ctggacgatg agaccgagaa gggcttcctg 1980

gagcgcaacc tgaatgatac aaggtacatc gcccggttca tgtgccagtt cgtgcaggag 2040

aacctgtatc tgaccggcaa gggcaagcgc ctggtgtttg cctccaacgg cggcatgacc 2100

gccacactgc ggaatctgtg gggcctgagg aaggtgcgcg aggataatga cagacaccac 2160

gcactggacg caatcgtggt ggcatgcagc accgcatcca tgcagcagaa gatcacaaag 2220

gcctttcagc ggcacgagag catcgagtat gtggataccg agacaggcga ggtgaagttc 2280

agaatccccc agccttggga cttctttcgc caggaagtga tgatccgggt gttttccgat 2340

cagccatgtg aggacctggt ggagaagctg tctgccaggc cagaggccct gcacgataac 2400

gtgacccctc tgttcgtgag cagggcacca aatagaaaga tgtccggcca gggccacctg 2460

gagacaatca agtccgccaa gcgcctgtcc gaggagaact ctatggtgaa gaagcccctg 2520

accacactga agctgaagga catccctgag atcgtgggct acccatctag agagccccag 2580

ctgtatgccg ccctgaagac caggctggag acacacgacg atgacccaat caaggccttt 2640

gccaagcctt tctacaagcc aaacaagaat ggagagctgg gcgccctggt gcggtccgtg 2700

agagtgaagg gcgtgcagaa cacaggcgtg atggtgcacg atggcaaggg catcgccgac 2760

aatgccacaa tggtgcgggt ggacgtgtat acaaaggccg gcaagaacta cctggtgccc 2820

gtgtacgtgt ggcaggtggc acagggaatc ctgccaaata gagccgtgac ctctggcaag 2880

agcgaggccg attgggacct gatcgatgag agcttcgagt ttaagttctc tctgagcaga 2940

ggcgacctgg tggagatgat ctccaacaag ggcaggatct tcggctacta taacggcctg 3000

gatagagcca atggcagcat cggcatcagg gagcacgatc tggagaagtc caagggcaag 3060

gacggagtgc acagggtggg agtgaagacc gcaacagcct ttaataagta ccacgtggac 3120

cccctgggca aggagatcca cagatgtagc tccgagccaa ggcccaccct gaagatcaag 3180

agcaagaagg gcaccggcgg gcccaagaag aagaggaggt atacccatac gatgttcctg 3240

actatgcggg ctatccctat gacgtcccgg actatgcagg atcgtatcct tat 3293

<210> 241
<211> 1121
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 241

Met Glu Lys Phe His Tyr Val Leu Gly Leu Asp Leu Gly Ile Ala Ser
1 5 10 15

Val Gly Trp Ala Ala Ile Glu Ile Asp Lys Glu Thr Glu Thr Ser Ile
20 25 30

Gly Leu Leu Asp Cys Gly Val Arg Thr Phe Glu Arg Ala Glu Val Pro
35 40 45

Lys Thr Gly Asp Ser Leu Ala Lys Ala Arg Arg Glu Ala Arg Ser Thr
50 55 60

```
Arg Arg Leu Ile Arg Arg Arg Ser His Arg Leu Leu Arg Leu Lys Arg
65                  70                  75                  80

Leu Leu Lys Arg Glu Ile Phe Arg Gln Pro Glu Thr Phe Lys Asp Leu
                85                  90                  95

Pro Ile Asn Ala Trp Gln Leu Arg Val Lys Gly Leu Asp Ser Arg Leu
            100                 105                 110

Asn Glu Tyr Glu Trp Ala Ala Val Leu Leu His Leu Val Lys His Arg
        115                 120                 125

Gly Tyr Leu Ser Gln Arg Lys Ser Glu Met Ser Glu Thr Asp Ser Lys
    130                 135                 140

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

Leu Gln Gln Glu Gln Tyr Arg Thr Pro Ala Glu Leu Ala Leu Lys Lys
                165                 170                 175

Phe Val Lys His Phe Arg Asn Lys Gly Gly Asp Tyr Ala His Thr Phe
            180                 185                 190

Asn Arg Leu Asp Leu Gln Ala Glu Leu His Leu Leu Phe Gln Lys Gln
        195                 200                 205

Arg Glu Leu Gly Asn Pro Phe Thr Ser Pro Glu Leu Glu Arg Gln Val
    210                 215                 220

Asp Asp Leu Leu Met Thr Gln Arg Ser Ala Leu Gln Gly Asp Ala Ile
225                 230                 235                 240

Leu Lys Met Leu Gly His Cys Gly Phe Glu Pro Glu Gln Phe Lys Ala
                245                 250                 255

Ala Lys Asn Thr Phe Ser Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu
            260                 265                 270

Asn Asn Leu Arg Ile Gln Asp Gln Gly Lys Glu Arg Ala Leu Thr Ala
        275                 280                 285

Asp Glu Arg Thr Lys Leu Leu Asp Glu Pro Tyr Lys Lys Ser Lys Leu
    290                 295                 300

Thr Tyr Ala Gln Val Arg Lys Leu Leu Ser Leu Pro Gln Thr Ala Ile
```

305 310 315 320

Phe Lys Gly Leu Arg Tyr Asp Leu Glu His Asp Lys Lys Ala Glu Asn
325 330 335

Ser Thr Leu Met Glu Met Lys Ser Tyr His Asn Ile Arg Gln Thr Leu
340 345 350

Glu Lys Ser Gly Leu Lys Thr Glu Trp Gln Ser Ile Ala Thr Gln Pro
355 360 365

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

Glu Asp Ile Ser His Glu Leu Lys Thr Cys Arg Leu Pro Glu Asn Val
385 390 395 400

Leu Asn Glu Leu Leu Lys Asn Ile Asn Phe Asp Gly Phe Ile Gln Leu
405 410 415

Ser Leu Thr Ala Leu Arg Lys Ile Leu Pro Leu Met Glu Gln Gly Tyr
420 425 430

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

Ser Leu Gln Gln Glu Ser Lys Gln Phe Leu Pro His Ile Pro Ile Asp
450 455 460

Asp Val Arg Asn Pro Val Val Phe Arg Thr Leu Thr Gln Ala Arg Lys
465 470 475 480

Val Val Asn Ala Ile Ile Arg Arg Tyr Gly Ser Pro Ala Arg Val His
485 490 495

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

Ile Glu Lys Gln Gln Gln Lys Asn Lys Lys Glu Arg Glu Asn Ala Val
515 520 525

Ala Lys Phe Lys Glu Asp Phe Pro Asp Phe Val Gly Glu Pro Arg Gly
530 535 540

Lys Asp Ile Leu Lys Met Arg Leu Tyr Glu Gln Gln His Gly Lys Cys
545 550 555 560

```
Leu Tyr Ser Gly His Asp Ile Asp Ile Asn Arg Leu Asn Glu Lys Gly
                565                 570                 575

Tyr Val Glu Ile Asp His Ala Leu Pro Phe Ser Arg Thr Trp Asp Asp
            580                 585                 590

Ser Gln Asn Asn Lys Val Leu Val Leu Gly Ser Glu Asn Gln Asn Lys
        595                 600                 605

Arg Asn Gln Thr Pro Asp Glu Tyr Leu Asp Gly Ala Asn Asn Ser Gln
    610                 615                 620

Arg Trp Leu Glu Phe Gln Ala Arg Val Gln Thr Cys His Phe Ser Tyr
625                 630                 635                 640

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

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

Phe Met Cys Gln Phe Val Gln Glu Asn Leu Tyr Leu Thr Gly Lys Gly
        675                 680                 685

Lys Arg Leu Val Phe Ala Ser Asn Gly Gly Met Thr Ala Thr Leu Arg
    690                 695                 700

Asn Leu Trp Gly Leu Arg Lys Val Arg Glu Asp Asn Asp Arg His His
705                 710                 715                 720

Ala Leu Asp Ala Ile Val Val Ala Cys Ser Thr Ala Ser Met Gln Gln
                725                 730                 735

Lys Ile Thr Lys Ala Phe Gln Arg His Glu Ser Ile Glu Tyr Val Asp
            740                 745                 750

Thr Glu Thr Gly Glu Val Lys Phe Arg Ile Pro Gln Pro Trp Asp Phe
        755                 760                 765

Phe Arg Gln Glu Val Met Ile Arg Val Phe Ser Asp Gln Pro Cys Glu
    770                 775                 780

Asp Leu Val Glu Lys Leu Ser Ala Arg Pro Glu Ala Leu His Asp Asn
785                 790                 795                 800
```

```
Val Thr Pro Leu Phe Val Ser Arg Ala Pro Asn Arg Lys Met Ser Gly
                805                 810                 815

Gln Gly His Leu Glu Thr Ile Lys Ser Ala Lys Arg Leu Ser Glu Glu
            820                 825                 830

Asn Ser Met Val Lys Lys Pro Leu Thr Ile Leu Lys Leu Lys Asp Ile
        835                 840                 845

Pro Glu Ile Val Gly Tyr Pro Ser Arg Glu Pro Gln Leu Tyr Ala Ala
    850                 855                 860

Leu Lys Thr Arg Leu Glu Thr His Asp Asp Asp Pro Ile Lys Ala Phe
865                 870                 875                 880

Ala Lys Pro Phe Tyr Lys Pro Asn Lys Asn Gly Glu Leu Gly Ala Leu
                885                 890                 895

Val Arg Ser Val Arg Val Lys Gly Val Gln Asn Thr Gly Val Met Val
            900                 905                 910

His Asp Gly Lys Gly Ile Ala Asp Asn Ala Thr Met Val Arg Val Asp
        915                 920                 925

Val Tyr Thr Lys Ala Gly Lys Asn Tyr Leu Val Pro Val Tyr Val Trp
    930                 935                 940

Gln Val Ala Gln Gly Ile Leu Pro Asn Arg Ala Val Thr Ser Gly Lys
945                 950                 955                 960

Ser Glu Ala Asp Trp Asp Leu Ile Asp Glu Ser Phe Glu Phe Lys Phe
                965                 970                 975

Ser Leu Ser Arg Gly Asp Leu Val Glu Met Ile Ser Asn Lys Gly Arg
            980                 985                 990

Ile Phe Gly Tyr Tyr Asn Gly Leu  Asp Arg Ala Asn Gly  Ser Ile Gly
        995                 1000                1005

Ile Arg  Glu His Asp Leu Glu  Lys Ser Lys Gly Lys  Asp Gly Val
    1010                1015                1020

His Arg  Val Gly Val Lys Thr  Ala Thr Ala Phe Asn  Lys Tyr His
    1025                1030                1035
```

Val Asp Pro Leu Gly Lys Glu Ile His Arg Cys Ser Ser Glu Pro
1040 1045 1050

Arg Pro Thr Leu Lys Ile Lys Ser Lys Lys Gly Thr Gly Gly Pro
1055 1060 1065

Lys Lys Lys Arg Lys Val Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1070 1075 1080

Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Gly Ser Tyr Pro Tyr
1085 1090 1095

Asp Val Pro Asp Tyr Ala Gly Ser Ala Ala Pro Ala Ala Lys Lys
1100 1105 1110

Lys Lys Leu Asp Phe Glu Ser Gly
1115 1120

<210> 242
<211> 148
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 242
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccuuuc ucauuucgga aacgaaauga 60

gaaccguugc uacaauaagg ccgucugaaa agaugugccg caacgcucug ccccuuaaag 120

cuucugcuuu aaggggcauc guuuauuu 148

<210> 243
<211> 141
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 243

nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccuuuu ucauuucgca gaaaugcgaa 60

augaaaaacg uuguuacaau aagagaaaag auuucucgca aagcucuguc ccuugaaaug 120

uaaguuucaa gggacaucuu u 141

<210> 244
<211> 172
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(24)
<223> n is a, c, g, t or u

<400> 244
nnnnnnnnnn nnnnnnnnnn nnnnguugua gcucccucuc ucaucucgua guggaaacac 60

uacgagauga gagccguugc uacaauaagg ccgucugaaa agacgcgccg cgacguaaaa 120

uacuuuaugu augagccccu guuugaguuu ucuuaaacag gggcaucguc uu 172

<210> 245
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 245
gtgaacttgt ggccgtttac gtcg 24

<210> 246
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 246
gtgaacttgt ggccgtttac gtcc 24

<210> 247
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 247
gtgaacttgt ggccgtttac gtgg 24

<210> 248
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 248
gtgaacttgt ggccgtttac gacg 24

<210> 249
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 249
gtgaacttgt ggccgtttac ctcg 24

<210> 250
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 250
gtgaacttgt ggccgtttag gtcg 24

<210> 251
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 251
gtgaacttgt ggccgttttc gtcg 24

<210> 252
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 252

gtgaacttgt ggccgttaac gtcg 24

<210> 253
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 253
gtgaacttgt ggccgtatac gtcg 24

<210> 254
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 254
gtgaacttgt ggccgattac gtcg 24

<210> 255
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 255
gtgaacttgt ggcccttac gtcg 24

<210> 256
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 256
gtgaacttgt ggcggtttac gtcg 24

<210> 257
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 257
gtgaacttgt gggcgtttac gtcg 24

<210> 258
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 258
gtgaacttgt gcccgtttac gtcg 24

<210> 259
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 259
gtgaacttgt cgccgtttac gtcg 24

<210> 260
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 260
gtgaacttga ggccgtttac gtcg 24

<210> 261
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 261
gtgaacttct ggccgtttac gtcg 24

<210> 262
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 262
gtgaactagt ggccgtttac gtcg 24

<210> 263
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 263
gtgaacatgt ggccgtttac gtcg 24

<210> 264
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 264
gtgaagttgt ggccgtttac gtcg 24

<210> 265
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 265
gtgatcttgt ggccgtttac gtcg 24

<210> 266
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 266
gtgtacttgt ggccgtttac gtcg 24

<210> 267
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 267
gtcaacttgt ggccgtttac gtcg 24

<210> 268
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 268
gagaacttgt ggccgtttac gtcg 24

<210> 269
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (1)..(3)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5)..(7)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (9)..(28)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (31)..(31)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (34)..(35)
<223> n is a, c, g, or t

<400> 269
nnncnnncnn nnnnnnnnnn nnnnnnnngg nccnn 35

<210> 270
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature

<222> (1)..(2)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5)..(5)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (8)..(27)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (29)..(31)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (33)..(35)
<223> n is a, c, g, or t

<400> 270
nnggnccnnn nnnnnnnnnn nnnnnnngnn ngnnn 35

<210> 271
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 271
tgaggaccgc cctgggcctg ggagaatccc t 31

<210> 272
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 272
gaaggaccac cctaggcctg ggagactccc t 31

<210> 273
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 273
ctcactcacc cacacagaca cacacgtcct 30

<210> 274
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 274
cacacacacc cacacagaca cacccccccc 30

<210> 275
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 275
ctccctcaca cacacagaca cacacctccc 30

<210> 276
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 276
cacacacaca cacacagaca cacacacccc 30

<210> 277
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 277
gtgtgtccct ctccccaccc gtccctgtcc g 31

<210> 278
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 278
ttgtctccct gtccccaccc gtccccttca g 31

<210> 279
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 279
ctgcctccct ctgcccaccc gtccttccca c 31

<210> 280
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 280
ctgtgcctct ctccccaccc ttccacaccc t 31

<210> 281
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 281
gctcatcccc ctccccaccc gtcctcgccc g 31

<210> 282
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 282
guuguagcuc ccgaaacguu gcuacaauaa ggccgucuga aaagaugugc cgcaacgcuc 60

ugccuucugg caucguu 77

<210> 283
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 283
ggagccccga aacggcacaa aaggccgcga aaagaggccg caacgccgcc cggcacga 58

<210> 284
<211> 109
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 284
atcctggtcg agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cggctttggc 60

gagacaaatc acctgcctgc tggaatacgg taaacctacg gcaagctga 109

<210> 285
<211> 109
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 285
tcagcttgcc gtaggtttac cgtattccac gaggcaggtg atttgtctcg ccaaagccgg 60

acacgctgaa cttgtggccg tttacgtcgc cgtccagctc gaccaggat 109

<210> 286
<211> 35
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 286

Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val
1 5 10 15

Ser Gly Phe Gly Glu Thr Asn His Leu Pro Arg Gly Ile Arg Thr Tyr
20 25 30

Gly Lys Leu
35

<210> 287
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 287

ccccaccggc gctggtgccc aggacgagga tggagattat gaag 44

<210> 288
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 288
cttcataatc tccatcctcg tcctgggcac cagcgccggt gggg 44

<210> 289
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 289
tgggagaatc ccttcccccc cttccctcgt gatctgcaac tcca 44

<210> 290
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 290
tggagttgca gatcacgagg gaagaggggg aagggattct ccca 44

<210> 291
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 291
gagaagatca actacactgg ccgtttctta cctggattcg aggc 44

<210> 292
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 292
gcctcgaatc caggtaagaa acggccagtg tagttgatct tctc 44

<210> 293
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 293
ccttctttgt ttcccctcct cccaggaata tgtggactat aatg 44

<210> 294
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 294
cattatagtc cacatattcc tgggaggagg ggaaacaaag aagg 44

<210> 295
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 295
cggcgctggt gcccaggacg aggatggaga tt 32

<210> 296
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 296
ggcgctggtg tcctggacga ggaactggac t 31

<210> 297
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 297
aggaactgga gcaaaggaca aggagatggt tt 32

<210> 298
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 298
agggaactgg gaccaggaca aggagcttga tt 32

<210> 299
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 299
aggtgcggga ggcgagggca agacttagtg ct 32

<210> 300
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 300
tgcagatcac gagggaagag ggggaaggga tt 32

<210> 301
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 301
gtcataacac cagggaagag gaggcctgga tt 32

<210> 302
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 302
tggcagaaca gagggaagag agggggaggg at 32

<210> 303
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 303
agcagatcag agggaagagg gaggggtgga ga 32

<210> 304
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 304
tacacatctc tgagggtgat gggcttgggg ct 32

<210> 305
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 305
tgcagaggtt tacaaacggg gggggggggg gg 32

<210> 306
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 306
gccagaagac caggtaggac ttggacgaca ag 32

<210> 307
<211> 1082
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 307

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 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 Thr Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asn
                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 Gly 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 Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255
```

```
His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
                405                 410                 415

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
```

```
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
        595                 600                 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
        675                 680                 685

Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
                725                 730                 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750
```

```
Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
        755                 760                 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        835                 840                 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
    850                 855                 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865                 870                 875                 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
                885                 890                 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
            900                 905                 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
        915                 920                 925

Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
    930                 935                 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
945                 950                 955                 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
                965                 970                 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
            980                 985                 990
```

```
Asp Ser Phe Asn Phe Lys Phe Ser  Leu His Pro Asn Asp  Leu Val Glu
        995                 1000                1005

Val Ile  Thr Lys Lys Ala Arg  Met Phe Gly Tyr Phe  Ala Ser Cys
    1010                1015                1020

His Arg  Gly Thr Gly Asn Ile  Asn Ile Arg Ile His  Asp Leu Asp
    1025                1030                1035

His Lys  Ile Gly Lys Asn Gly  Ile Leu Glu Gly Ile  Gly Val Lys
    1040                1045                1050

Thr Ala  Leu Ser Phe Gln Lys  Tyr Gln Ile Asp Glu  Leu Gly Lys
    1055                1060                1065

Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg Pro Pro  Val Arg
    1070                1075                1080
```

<210> 308
<211> 1082
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 308

```
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 Asp Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
    210                 215                 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225                 230                 235                 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
                245                 250                 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
            260                 265                 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
        275                 280                 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
    290                 295                 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305                 310                 315                 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
                325                 330                 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
            340                 345                 350
```

```
Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Ser Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

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

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Val
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590
```

```
Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
        595                 600                 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Cys Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp His Ile Leu Leu Thr
        675                 680                 685

Gly Lys Gly Lys Arg Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
                725                 730                 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Lys Val Leu His Gln
        755                 760                 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Ala His Lys Asp Thr Leu Arg Ser
```

835 840 845

Ala Lys Arg Phe Val Lys His Asn Glu Lys Ile Ser Val Lys Arg Val
850 855 860

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

Lys Asn Gly Arg Glu Ile Glu Leu Tyr Glu Ala Leu Lys Ala Arg Leu
885 890 895

Glu Ala Tyr Gly Gly Asn Ala Lys Gln Ala Phe Asp Pro Lys Asp Asn
900 905 910

Pro Phe Tyr Lys Lys Gly Gly Gln Leu Val Lys Ala Val Arg Val Glu
915 920 925

Lys Thr Gln Glu Ser Gly Val Leu Leu Asn Lys Lys Asn Ala Tyr Thr
930 935 940

Ile Ala Asp Asn Gly Asp Met Val Arg Val Asp Val Phe Cys Lys Val
945 950 955 960

Asp Lys Lys Gly Lys Asn Gln Tyr Phe Ile Val Pro Ile Tyr Ala Trp
965 970 975

Gln Val Ala Glu Asn Ile Leu Pro Asp Ile Asp Cys Lys Gly Tyr Arg
980 985 990

Ile Asp Asp Ser Tyr Thr Phe Cys Phe Ser Leu His Lys Tyr Asp Leu
995 1000 1005

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

Tyr Ile Asn Cys Asp Ser Ser Asn Gly Arg Phe Tyr Leu Ala Trp
1025 1030 1035

His Asp Lys Gly Ser Lys Glu Gln Gln Phe Arg Ile Ser Thr Gln
1040 1045 1050

Asn Leu Val Leu Ile Gln Lys Tyr Gln Val Asn Glu Leu Gly Lys
1055 1060 1065

Glu Ile Arg Pro Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1070 1075 1080

<210> 309
<211> 1082
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 309

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 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 Thr Arg Arg Leu Leu Lys Arg Glu Gly Val Leu Gln Ala Ala Asn
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 Gly 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 Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu

195 200 205

Gln Ala Glu Leu Ile Leu Leu Phe Glu Lys Gln Lys Glu Phe Gly Asn
210 215 220

Pro His Val Ser Gly Gly Leu Lys Glu Gly Ile Glu Thr Leu Leu Met
225 230 235 240

Thr Gln Arg Pro Ala Leu Ser Gly Asp Ala Val Gln Lys Met Leu Gly
245 250 255

His Cys Thr Phe Glu Pro Ala Glu Pro Lys Ala Ala Lys Asn Thr Tyr
260 265 270

Thr Ala Glu Arg Phe Ile Trp Leu Thr Lys Leu Asn Asn Leu Arg Ile
275 280 285

Leu Glu Gln Gly Ser Glu Arg Pro Leu Thr Asp Thr Glu Arg Ala Thr
290 295 300

Leu Met Asp Glu Pro Tyr Arg Lys Ser Lys Leu Thr Tyr Ala Gln Ala
305 310 315 320

Arg Lys Leu Leu Gly Leu Glu Asp Thr Ala Phe Phe Lys Gly Leu Arg
325 330 335

Tyr Gly Lys Asp Asn Ala Glu Ala Ser Thr Leu Met Glu Met Lys Ala
340 345 350

Tyr His Ala Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
355 360 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
370 375 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385 390 395 400

Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
405 410 415

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
420 425 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
435 440 445

```
Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
    530                 535                 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545                 550                 555                 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
                565                 570                 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
            580                 585                 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
        595                 600                 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
        675                 680                 685
```

```
Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
                725                 730                 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
        755                 760                 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780

Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Leu Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        835                 840                 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
    850                 855                 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865                 870                 875                 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
                885                 890                 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
            900                 905                 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
        915                 920                 925
```

```
Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
    930                 935                 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
945                 950                 955                 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
                965                 970                 975

Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
            980                 985                 990

Asp Ser Phe Asn Phe Lys Phe Ser  Leu His Pro Asn Asp  Leu Val Glu
        995                 1000                1005

Val Ile  Thr Lys Lys Ala Arg  Met Phe Gly Tyr Phe  Ala Ser Cys
    1010                1015                1020

His Arg  Gly Thr Gly Asn Ile  Asn Ile Arg Ile His  Asp Leu Asp
    1025                1030                1035

His Lys  Ile Gly Lys Asn Gly  Ile Leu Glu Gly Ile  Gly Val Lys
    1040                1045                1050

Thr Ala  Leu Ser Phe Gln Lys  Tyr Gln Ile Asp Glu  Leu Gly Lys
    1055                1060                1065

Glu Ile  Arg Pro Cys Arg Leu  Lys Lys Arg Pro Pro  Val Arg
    1070                1075                1080
```

<210> 310
<211> 1081
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 310

```
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 Asn
                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 Gly 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 Cys Gly His Ile
            180                 185                 190

Arg Asn Gln Arg Ser Asp Tyr Ser His Thr Phe Ser Arg Lys Asp Leu
        195                 200                 205

Gln Ala Glu Leu Asn 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 Ser 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 Thr Ile Ser Arg Ala Leu Glu Lys Glu Gly Leu Lys Asp Lys
        355                 360                 365

Lys Ser Pro Leu Asn Leu Ser Pro Glu Leu Gln Asp Glu Ile Gly Thr
    370                 375                 380

Ala Phe Ser Leu Phe Lys Thr Asp Glu Asp Ile Thr Gly Arg Leu Lys
385                 390                 395                 400

Asp Arg Ile Gln Pro Glu Ile Leu Glu Ala Leu Leu Lys His Ile Ser
                405                 410                 415

Phe Asp Lys Phe Val Gln Ile Ser Leu Lys Ala Leu Arg Arg Ile Val
            420                 425                 430

Pro Leu Met Glu Gln Gly Lys Arg Tyr Asp Glu Ala Cys Ala Glu Ile
        435                 440                 445

Tyr Gly Asp His Tyr Gly Lys Lys Asn Thr Glu Glu Lys Ile Tyr Leu
    450                 455                 460

Pro Pro Ile Pro Ala Asp Glu Ile Arg Asn Pro Val Val Leu Arg Ala
465                 470                 475                 480

Leu Ser Gln Ala Arg Lys Val Ile Asn Gly Val Val Arg Arg Tyr Gly
                485                 490                 495

Ser Pro Ala Arg Ile His Ile Glu Thr Ala Arg Glu Val Gly Lys Ser
            500                 505                 510

Phe Lys Asp Arg Lys Glu Ile Glu Lys Arg Gln Glu Glu Asn Arg Lys
        515                 520                 525

Asp Arg Glu Lys Ala Ala Ala Lys Phe Arg Glu Tyr Phe Pro Asn Phe
```

530 535 540

Val Gly Glu Pro Lys Ser Lys Asp Ile Leu Lys Leu Arg Leu Tyr Glu
545 550 555 560

Gln Gln His Gly Lys Cys Leu Tyr Ser Gly Lys Glu Ile Asn Leu Gly
565 570 575

Arg Leu Asn Glu Lys Gly Tyr Val Glu Ile Asp His Ala Leu Pro Phe
580 585 590

Ser Arg Thr Trp Asp Asp Ser Phe Asn Asn Lys Val Leu Val Leu Gly
595 600 605

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

Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625 630 635 640

Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
645 650 655

Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
660 665 670

Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
675 680 685

Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
690 695 700

Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705 710 715 720

Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
725 730 735

Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
740 745 750

Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
755 760 765

Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
770 775 780

```
Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800

Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815

Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830

Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        835                 840                 845

Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
    850                 855                 860

Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865                 870                 875                 880

Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
                885                 890                 895

Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
            900                 905                 910

Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
        915                 920                 925

Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
    930                 935                 940

Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
945                 950                 955                 960

Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
                965                 970                 975

Arg Ala Val Val Ala Tyr Ala Asp Glu Glu Asp Trp Thr Val Ile Asp
            980                 985                 990

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

Val Gln  Leu Lys Lys Asp Ser  Phe Leu Gly Tyr Phe  Ser Gly Leu
    1010                1015                1020
```

```
Asp Arg  Ala Thr Gly Ala Ile  Ser Leu Arg Glu His  Asp Leu Glu
    1025                1030                1035

Lys Ser  Lys Gly Lys Asp Gly  Met His Arg Ile Gly  Val Lys Thr
    1040                1045                1050

Ala Leu  Ser Phe Gln Lys Tyr  Gln Ile Asp Glu Met  Gly Lys Glu
    1055                1060                1065

Ile Arg  Pro Cys Arg Leu Lys  Lys Arg Pro Pro Val  Arg
    1070                1075                1080
```

<210> 311
<211> 1140
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 311

```
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> 312
<211> 1168
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 312

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> 313
<211> 1135
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic

<400> 313

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> 314
<211> 4377
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (137)..(159)
<223> n is a, c, g, or t

<400> 314
gtttaaacaa aaaaataaac gatgcccctt aaagcagaag ctttaagggg cagagcgttg 60

cggcacatct tttcagacgg ccttattgta gcaacggttc tcatttcgtt tccgaaatga 120

gaaagggagc tacaacnnnn nnnnnnnnnn nnnnnnnnnc ggtgtttcgt cctttccaca 180

agatatataa agccaagaaa tcgaaatact ttcaagttac ggtaagcata tgatagtcca 240

ttttaaaaca taattttaaa actgcaaact acccaagaaa ttattacttt ctacgtcacg 300

tattttgtac taatatcttt gtgtttacag tcaaattaat tctaattatc tctctaacag 360

ccttgtatcg tatatgcaaa tatgaaggaa tcatgggaaa taggccctct tcctgcccga 420

ccttgacgtc gactctagaa tggaggcggt actatgtaga tgagaattca ggagcaaact 480

gggaaaagca actgcttcca aatatttgtg atttttacag tgtagttttg gaaaaactct 540

tagcctacca attcttctaa gtgttttaaa atgtgggagc cagtacacat gaagttatag 600

agtgttttaa tgaggcttaa atatttaccg taactatgaa atgctacgca tatcatgctg 660

ttcaggctcc gtggccacgc aactcatact taagcagaca gtggttcaaa gttttttttct 720

tccatttcag gtgtcgtgaa caccgccacc atggtgccta agaagaagag aaaggtggaa 780

gataaacgcc cagcagctac aaagaaggca ggtcaagcca agaaaaagaa agcagcattc 840

aagccaaact caatcaatta catcctggga ctggacatcg gcatcgcatc cgtcgggtgg 900

gctatggtcg aaatcgacga ggaggagaac cccatccgcc tgatcgatct gggcgtgcgc 960

gtgtttgaga gggcagaggt gcctaagacc ggcgacagcc tggccatggc acggagactg 1020

gcacgctccg tgaggcgcct gacccggaga agggcccaca gactgctgag gacacgccgg 1080

ctgctgaaga gggagggcgt gctgcaggcc gccaacttcg atgagaatgg cctgatcaag 1140

tccctgccca ataccccttg gcagctgagg gcagccgccc tggaccgcaa gctgacacct 1200

ctggagtggt ccgccgtgct gctgcacctg atcaagcacc ggggctacct gtctcagaga 1260

aagaacgagg gcgagacagc cgataaggag ctgggcgccc tgctgaaggg agtggcagga 1320

aatgcacacg ccctgcagac cggcgacttt cgcacaccag ccgagctggc cctgaacaag 1380

ttcgagaagg agagcggcca catccgcaat cagcggtctg actatagcca caccttctcc 1440

cggaaggatc tgcaggccga gctgatcctg ctgtttgaga agcagaagga gttcggcaac 1500

ccacacgtgt ctggcggcct gaaggagggc atcgagacac tgctgatgac acagcggccc 1560

gccctgagcg gcgacgcagt gcagaagatg ctgggacact gcacctttga gccagccgag 1620

cccaaggccg ccaagaatac ctacacagcc gagcggttca tctggctgac aaagctgaac 1680

aatctgagga tcctggagca gggaagcgag cgcccactga ccgacacaga gagggccacc 1740

ctgatggatg agccctaccg caagtccaag ctgacatatg cacaggcaag gaagctgctg 1800

ggcctggagg acaccgcctt ctttaagggc ctgagatacg gcaaggataa cgccgaggcc 1860

tctacactga tggagatgaa ggcctatcac gccatcagca gggccctgga gaaggagggc 1920

ctgaaggaca agaagtcccc actgaatctg tctcccgagc tgcaggatga gatcggcacc 1980

gcctttagcc tgttcaagac cgacgaggat atcacaggca gactgaagga caggatccag 2040

ccagagatcc tggaggccct gctgaagcac atcagctttg ataagttcgt gcagatcagc 2100

ctgaaggccc tgcggaggat cgtgccactg atggagcagg gcaagaggta cgacgaggcc 2160

tgcgccgaaa tctacggcga tcactatggc aagaagaaca cagaggagaa aatctacctg 2220

ccccctatcc ccgccgatga gatcaggaac cctgtggtgc tgcgcgccct gtctcaggca 2280

agaaaagtga tcaacggagt ggtgcgccgg tacggcagcc ccgccagaat ccacatcgag 2340

acagccaggg aagtgggcaa gtcctttaag gacagaaagg agatcgagaa gaggcaggag 2400

gagaacagaa aggataggga gaaggccgcc gccaagttca gagagtactt tcctaatttc 2460

gtgggcgagc caaagtccaa ggatatcctg aagctgaggc tgtacgagca gcagcacggc 2520

aagtgtctgt attctggcaa ggagatcaac ctgggccgcc tgaatgagaa gggctatgtg 2580

gagatcgacc acgccctgcc tttttctcgg acctgggacg atagcttcaa caataaggtg 2640

ctggtgctgg gctctgagaa ccagaataag ggcaaccaga caccctacga gtatttcaac 2700

ggcaaggaca atagccgcga gtggcaggag tttaaggcaa gggtggagac aagcaggttc 2760

cctcggtcca agaagcagag aatcctgctg cagaagtttg acgaggatgg cttcaaggag 2820

aggaacctga atgacacccg ctacgtgaat cggtttctgt gccagttcgt ggccgataga 2880

atgaggctga ccggcaaggg caagaagaga gtgtttgcct ccaacggcca gatcacaaat 2940

ctgctgaggg gcttctgggg cctgagaaag gtgagggcag agaacgacag gcaccacgca 3000

ctggatgcag tggtggtggc atgttctacc gtggccatgc agcagaagat cacacgcttt 3060

gtgcggtata aggagatgaa tgccttcgac ggcaagacca tcgataagga gacaggcgag 3120

gtgctgcacc agaagacaca ctttcctcag ccatgggagt tctttgccca ggaagtgatg 3180

atccgggtgt ttggcaagcc tgacggcaag ccagagttcg aggaggccga taccctggag 3240

aagctgagaa cactgctggc agagaagctg agctccaggc ccgaggcagt gcacgagtac 3300

gtgaccccac tgttcgtgtc tagagccccc aacaggaaga tgagcggcca gggccacatg 3360

gagacagtga agtccgccaa gagactggac gagggcgtgt ctgtgctgag ggtgcctctg 3420

acacagctga agctgaagga tctggagaag atggtgaatc gcgagcggga gccaaagctg 3480

tatgaggccc tgaaggcaag gctggaggca cacaaggacg atcctgccaa ggcctttgcc 3540

gagccattct acaagtatga taaggccggc aacagaaccc agcaggtgaa ggccgtgagg 3600

gtggagcagg tgcagaagac aggcgtgtgg gtgcgcaacc acaatggcat cgccgacaat 3660

gctaccatgg tgcgggtgga cgtgtttgag aagggcgata agtactatct ggtgcccatc 3720

tacagctggc aggtggccaa gggcatcctg cctgatagag ccgtggtgca gggcaaggac 3780

gaggaggatt ggcagctgat cgacgattcc ttcaacttta agttctctct gcaccccaat 3840

gacctggtgg aagtgatcac caagaaggcc aggatgtttg gctacttcgc ctcctgccac 3900

cgcggcacag gcaacatcaa tatccggatc cacgacctgg atcacaagat cggcaagaac 3960

ggcatcctgg agggcatcgg cgtgaagaca gccctgagct tccagaagta tcagatcgac 4020

gagctgggca aggagatcag accttgtagg ctgaagaagc gcccacccgt gcgggaggat 4080

aagcggcccg cagcaaccaa gaaggcagga caggccaaga agaagaagta cccctatgac 4140

gtgcctgact acgccgggta tccctacgac gtgcctgatt acgccgggtc ctatccctac 4200

gacgtgccag attacgctgc agctccagca gcgaagaaaa agaagctgga ttaagatctt 4260

tttccctctg ccaaaaatta tggggacatc atgaagcccc ttgagcatct gacttctggc 4320

taataaagga aatttatttt cattgcaata gtgtgttgga atttttttgtg tctctca 4377

Claims

We claim:

1. A single guide ribonucleic acid (sgRNA) sequence comprising a truncated repeat:antirepeat region.

2. The sgRNA sequence of Claim 1, further comprising a truncated Stem 2 region.

3. The sgRNA sequence of Claim 2, further comprising a truncated spacer region.

4. The sgRNA sequence of Claim 1, wherein said sgRNA sequence has a length of 121 nucleotides.

5. The sgRNA sequence of Claim 2, wherein said sgRNA sequence length is selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides.

6. The sgRNA sequence of Claim 3, wherein said sgRNA sequence has a length of 100 nucleotides.

7. The sgRNA sequence of Claim 1, wherein said sgRNA sequence is an Nme1Cas9 single guide ribonucleic acid sequence or an Nme2Cas9 single guide ribonucleic acid sequence.

8. A single guide ribonucleic acid (sgRNA) sequence comprising a truncated Stem 2 region.

9. The sgRNA sequence of Claim 8, further comprising a truncated repeat:antirepeat region.

10. The sgRNA sequence of Claim 9, further comprising a truncated spacer region.

11. The sgRNA sequence of Claim 9, wherein said sgRNA sequence length is selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides.

12. The sgRNA sequence of Claim 10, wherein said sgRNA sequence has a length of 100 nucleotides.

13. An adeno-associated viral (AAV) plasmid comprising a single guide ribonucleic acid-*Neisseria meningitidis* Cas9 nucleic acid vector.

14. The AAV plasmid of Claim 13, wherein said single guide ribonucleic acid-*Neisseria meningitidis* Cas9 nucleic acid vector comprises at least one promoter.

15. The AAV plasmid of Claim 14, wherein said at least one promoter is selected from the group consisting of a U6 promoter and a U1a promoter.

16. The AAV plasmid of Claim 13, wherein said single guide ribonucleic acid-*Neisseria meningitidis* Cas9 nucleic acid vector comprises a Kozak sequence.

17. The AAV plasmid of Claim 13, wherein said sgRNA comprises a nucleic acid sequence that is complementary to a gene-of-interest sequence.

18. The AAV plasmid of Claim 17, wherein said gene-of-interest sequence is selected from the group consisting of a *PCSK9* sequence and a *ROSA26* sequence.

19. The AAV plasmid of Claim 13, wherein said sgRNA comprises a truncated repeat-antirepeat sequence.

20. The AAV plasmid of Claim 19, wherein said sgRNA further comprises a truncated Stem 2 region.

21. The AAV plasmid of Claim 20, wherein said sgRNA further comprises a truncated spacer region.

22. The AAV plasmid of Claim 19, wherein said sgRNA sequence has a length of 121 nucleotides.

23. The AAV plasmid of Claim 20, wherein said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides.

24. The AAV plasmid of Claim 21, wherein said sgRNA sequence has a length of 100 nucleotides.

25. The AAV plasmid of Claim 13, wherein said sgRNA comprises a truncated Stem 2 region.

26. The AAV plasmid of Claim 25, wherein said sgRNA further comprises a truncated repeat:antirepeat region.

27. The AAV plasmid of Claim 26, wherein said sgRNA further comprises a truncated spacer region.

28. The AAV plasmid of Claim 26, wherein said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides.

29. The AAV plasmid of Claim 27, wherein said sgRNA sequence has a length of 100 nucleotides.

30. A method, comprising:

a) providing;

i) a patient exhibiting at least one symptom of a medical condition, wherein said patient comprises a plurality of genes related to said medical condition;

ii) an adeno-associated viral (AAV) plasmid comprising a single guide ribonucleic acid-*Neisseria meningitidis* Cas9 nucleic acid vector, wherein said sgRNA comprises a nucleic acid sequence that is complementary to a portion of at least one of said plurality of genes; and

b) administering said AAV plasmid to said patient under conditions such that said at least one symptom of said medical condition is reduced.

31. The method of Claim 30, wherein said medical condition comprises hypercholesterolemia.

32. The method of Claim 30, wherein said at least one of said plurality of genes is a *PCSK9* gene.

33. The method of Claim 32, wherein said sgRNA nucleic acid is complementary to a portion of said *PCSK9* gene.

34. The method of Claim 30, wherein said sgRNA comprises a truncated repeat-antirepeat sequence.

35. The method of Claim 34, wherein said sgRNA further comprises a truncated Stem 2 region.

36. The method of Claim 35, wherein said sgRNA further comprises a truncated spacer region.

37. The method of Claim 34, wherein said sgRNA sequence has a length of 121 nucleotides.

38. The method of Claim 35, wherein said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides.

39. The method of Claim 21, wherein said sgRNA sequence has a length of 100 nucleotides.

40. The method of Claim 30, wherein said sgRNA comprises a truncated Stem 2 region.

41. The method of Claim 40, wherein said sgRNA further comprises a truncated repeat:antirepeat region.

42. The method of Claim 41, wherein said sgRNA further comprises a truncated spacer region.

43. The method of Claim 41, wherein said sgRNA sequence has a length selected from the group consisting of 111 nucleotides, 107 nucleotides, 105 nucleotides, 103 nucleotides, 102 nucleotides, 101 nucleotides, and 99 nucleotides.

44. The method of Claim 42, wherein said sgRNA sequence has a length of 100 nucleotides.

45. An adeno-associated viral (AAV) plasmid encoding a Type II-C Cas9 nuclease protein wherein said protein comprises a protospacer adjacent motif recognition domain configured with a bind site to a protospacer adjacent motif sequence comprising between one – four required nucleotides.

46. The adeno-associated viral plasmid of Claim 45, wherein said Type II-C Cas9 nuclease protein is selected from the group consisting of a *Neisseria meningitidis strain De10444* Nme2Cas9 nuclease protein, a *Haemophilus parainfluenzae* HpaCas9 nuclease protein and a *Simonsiella muelleri* SmuCas9 nuclease protein.

47. The adeno-associated viral plasmid of Claim 46, wherein said protospacer adjacent motif sequence comprising between one - four required nucleotides is selected from the group consisting of N_4CN_3, N_4CT, N_4CCN, N_4CCA, and N_4GNT_3.

48. The adeno-assocaited viral plasmid of Claim 45, wherein said one – four required nucleotides is selected from the group consisting of C, CC, CT, CCN, CCA, CN_3 and GNT_2.

49. The adeno-associated viral plasmid of Claim 45, wherein said Type II-C Cas9 nuclease protein is bound to a truncated sgRNA.

50. A method, comprising:

a) providing;

i) a patient exhibiting at least one symptom of a medical condition, wherein said patient comprises a plurality of genes related to said medical condition, wherein said plurality of genes comprise a protospacer adjacent motif comprising between two – four required nucleotides;

ii) a delivery platform comprising at least one nucleic acid encoding a Type II-C Cas9 nuclease protein wherein said protein comprises a protospacer adjacent motif recognition domain configured with a bind site to said protospacer adjacent motif sequence comprising between two – four required nucleotides; and

b) administering said delivery platform to said patient under conditions such that said at least one symptom of said medical condition is reduced.

51. The method of Claim 50, wherein said delivery platform comprises an adeno-associated viral plasmid.

52. The method of Claim 50, wherein said delivery platform comprises a microparticle.

53. The method of Claim 50, wherein said Type II-C Cas9 nuclease protein is selected from the group consisting of a *Neisseria meningitidis strain De10444* Nme2Cas9 nuclease protein, a *Haemophilus parainfluenzae* HpaCas9 nuclease protein and a *Simonsiella muelleri* SmuCas9 nuclease protein.

54. The method of Claim 50, wherein said protospacer adjacent motif sequence comprising one - four required nucleotides is selected from the group consisting of N_4C_3, N_4CT, N_4CCN, N_4CCA, and N_4GNT_3.

55. The adeno-associated viral plasmid of Claim 50, wherein said one – four required nucleotides are selected from the group consisting of C, CC, CT, CCN, CCA, CN_3 and GNT_2.

56. The method of Claim 50, wherein said Type II-C Cas9 nuclease protein is bound to a truncated sgRNA.

57. The method of Claim 50, wherein said medical condition is selected from the group consisting of hyperlipidemia and tyrosinemia.