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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
04 November 2021 (04.11.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/222899 A1

(51) International Patent Classification:
C07K 14/005 (2006.01) *C12N 15/86* (2006.01)
C07K 14/015 (2006.01)

(21) International Application Number:
PCT/US2021/030489

(22) International Filing Date:
03 May 2021 (03.05.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/018,573 01 May 2020 (01.05.2020) US

(71) Applicant: **TRUSTEES OF BOSTON COLLEGE** [US/US]; 140 Commonwealth Avenue, Chestnut Hill, Massachusetts 02467 (US).

(72) Inventors: **CHATTERJEE, Abhishek**; 22 Chestnut Pl, #304, Brookline, Massachusetts 02445 (US). **KELEMEN, Rachel E.**; 9 Payne Road, Newton, Massachusetts 02461 (US). **ERICKSON, Sarah B.**; 48 Burbank Street, Apt. 8, Boston, Massachusetts 02115 (US). **PHAM, Quan**; 78 Beacon Street, Chestnut Hill, Massachusetts 02467 (US).

(74) Agent: **HOGLE, Doreen M.**; HOUSTONHOGLE LLP, 1666 Massachusetts Avenue, Suite 12, Lexington, Massachusetts 02420 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

(54) Title: CONTROLLED MODIFICATION OF ADENO-ASSOCIATED VIRUS (AAV) FOR ENHANCED GENE THERAPY

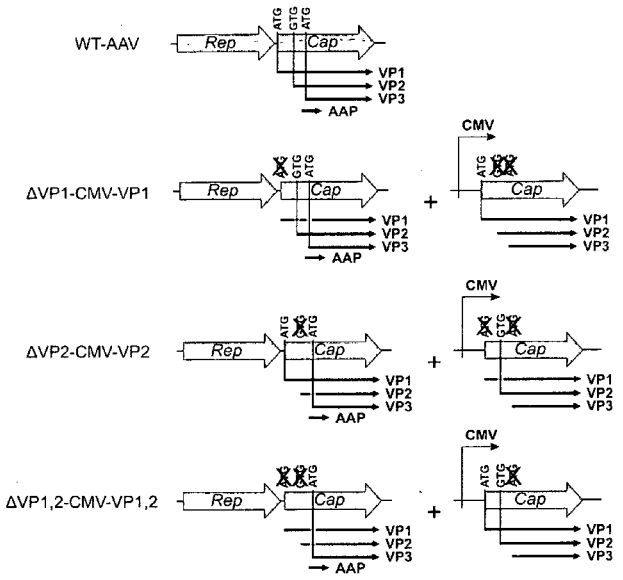


Figure 2

(57) Abstract: The present invention discloses platforms for chemically modify AAV capsids with control over site and stoichiometry. An AAV packaging system is described that allows the introduction of site-directed natural and unnatural amino acid mutations into any subset of the three capsid proteins. These engineered residues can be subsequently used to chemically functionalize the resulting capsids with precise control over site and stoichiometry. Such controlled modification strategy can be used to attach a wide variety of entities to AAV capsids to engineer its tropism, immunogenicity, etc.

[Continued on next page]

WO 2021/222899 A1

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

Published:

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

CONTROLLED MODIFICATION OF ADENO-ASSOCIATED VIRUS (AAV) FOR ENHANCED GENE THERAPY

GOVERNMENT SUPPORT

[0001] This invention was made with Government support under grant number 1817893, awarded by the National Science Foundation. The Government has certain rights in the invention.

RELATED APPLICATIONS

[0002] This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 63/018,573, filed on May 1, 2020, which is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE OF MATERIAL IN ASCII TEXT FILE

[0003] This application incorporates by reference the Sequence Listing contained in the following ASCII text file: : 0342_0010WO1_SL.txt; created May 3, 2021, 76,879 bytes in size.

FIELD OF THE INVENTION

[0004] The present invention is directed to the field of biotechnology, focusing on the development of engineered AAV-based gene-therapy vectors through the controlled chemical modification of the virus capsid.

BACKGROUND OF THE INVENTION

[0005] Adeno-Associated Virus (AAV) has emerged as one of the most promising delivery vehicles for gene therapy. It is naturally replication deficient, exhibits only a low innate immune response, efficiently infects dividing as well as non-dividing cells, and provides stable, long-term expression of the delivered therapeutic gene in vivo. Indeed, several AAV-based gene therapies have recently been approved, and many others are currently in clinical trial. Despite the promise of AAV-based gene therapy to provide cures for numerous diseases that are currently untreatable, there are several key challenges. For example, nearly all of the current approaches rely on the natural serotypes of AAV, which

exhibit restrictive innate tissue-tropism. Naturally, diseases associated with cell-types that are efficiently infected by these vectors can be targeted by this approach. However, many other diseases that occur in tissues that are not readily infected by existing AAV vectors currently remain refractory. The ability to reengineer the tissue-tropism of existing AAV vectors will be highly useful for creating custom vectors that could deliver genetic payloads to a desired location, substantially expanding the scope of AAV-based gene therapy. Additionally, many patients have pre-existing immunity to existing AAV vectors. Furthermore, administration of therapeutic AAV vectors elicits strong adaptive immune response, prohibiting a second dose of the same vector. The ability to engineer existing AAV vectors to evade such immune-recognition will be key for the development of sustainable gene therapies. Directed evolution has been used to develop AAV capsids with more therapeutically desirable traits. However, this approach is often unreliable for producing acceptable solutions, and the resulting mutant AAV capsids frequently show packaging deficiencies.

[0006] The ability to display activity-modulating entities on existing AAV capsids in a controlled manner provides an attractive modular approach to engineer their properties. Numerous methods have been attempted towards this goal, which include N-terminal extension of the minor capsid proteins, insertion of peptide loops in permissive sites of the capsid proteins, and modification of existing amino acids, such as lysine, on the capsid surface. However, N-terminal extension and loop insertion often disrupt the assembly of the complex AAV capsid. Because of the delicacy of the AAV capsid, insertion of peptide loops is restricted to very few locations and the resulting capsids often exhibit suboptimal packaging. These methods provide limited flexibility in optimizing the placement of the activity-modulating moiety, the structural variety of entities that can be displayed, as well as the ability to optimize the number of targeting moieties displayed on the capsid surface.

SUMMARY OF THE INVENTION

[0007] It has been possible to produce AAV incorporating UAAs with an azido functionality, which was then conjugated to a cyclooctyne-modified cRGDFC via strain-promoted azide-alkyne cycloaddition (SPAAC), thus retargeting the virus to the αVβ3 integrin receptor that is overexpressed in certain types of cancer cells. However, incorporation of UAAs as previously described has been restricted to the incorporation of the UAA to all 60 capsid proteins of AAV. As described herein, it has been demonstrated

that the over-modification of all 60 capsid proteins (including minor capsid proteins VP1 and VP2 as well as the major capsid protein VP3) of the AAV virus particle is strongly detrimental for its infectivity. The ability to control the number of amino acid mutations to produce variant capsids to incorporate/introduce a specific number of naturally-occurring and/or UAA modifications/mutations per capsid (relative to the wild-type AAV) is a key feature of the present invention to allow the synthesis of infectious genetically-modified adeno-associated viruses comprising variant capsid proteins with mutated amino acid residues. These mutated residues in the variant capsid proteins permit the attachment (e.g., covalent attachment) of biological and/or chemical entities (also referred to herein as "groups") to introduce bioconjugation "handles" into the AAV in a site-specific manner resulting in the incorporation/introduction of a defined number of activity-modulating groups that enable optimally engineering AAV function.

[0008] The present invention encompasses genetically engineered/genetically-modified adeno-associated virus (AAV) with modification/mutations at site specific locations in one, or more, of the capsid proteins relative to the corresponding sites of the wild-type AAV, and methods of making these genetically-modified viruses. Importantly, as described herein, modifications can be limited to one, or more specific, chemo-selective location(s) in a single capsid protein such as VP1, VP2 or VP3, or multiples of these capsid proteins, such as in both VP1 and VP2, without affecting the infectivity of the AAV particle (i.e., wherein the infectivity of the genetically-modified AAV is essentially comparable to, or slightly altered from, the infectivity of wild-type AAV under similar biological conditions).

[0009] AAV is a small non-enveloped parvovirus with a 4.5 kb single stranded DNA genome which codes for the non-structural Rep and AAP proteins and the structural Cap proteins VP1, VP2, and VP3, which share a common C terminus and differ by N-terminal extensions. Unlike enveloped viruses, which have a lipid membrane decorated with surface proteins, AAV has tightly packed icosahedral capsid consisting only of 60 copies of the proteins VP1, VP2, and VP3 in a roughly 1:1:10 stoichiometry. AAV is replication-deficient – complete virus production and escape requires the presence of a helper virus such as adenovirus. AAV has a number of different serotypes with different natural tropisms. For example, one serotype described herein is AAV type 2 (AAV2), which targets cells expressing the heparan sulfate proteoglycan receptor.

[0010] Adeno-associated virus has emerged as the most promising candidate for human gene therapy. This virus boasts desirable properties like low immunogenicity and long-term gene expression, ideal for a gene therapy vector. However, many of the other properties of these virus often do not align with therapeutic needs. For example, natural serotypes of the virus have distinct cell specificities, which restrict which cells/tissues can be targeted for human gene therapy. The ability to functionalize existing AAV vectors, with a precise control over site and stoichiometry offers an attractive avenue to introduce therapeutically desirable traits such as cell-specificity and immune-evasion.

[0011] Described herein is technology to precisely, and site-specifically, genetically-engineer AAV by introducing amino acids, such as UAAs, with entities or groups such as bioconjugation entities or "handles" (e.g., the natural amino acid cysteine, or unnatural amino acids with a broad range of bioconjugation chemistries) into the minor capsid proteins of AAV. The minor capsid proteins VP1 and VP2 are present at 5 copies each per AAV capsid. As described herein, the constructs and methods of the present invention enable site-selective modification of any of the three capsid proteins, but specifically modify VP1 or VP2 or both VP1 and VP2, allowing the attachment of 5 or 10 groups per fully-assembled AAV capsid. (the term "fully-assembled" as used herein means virus assembly incorporating all three capsid proteins, the minor VP1 and VP2 capsid proteins as well as the major capsid proteins, VP3 to form an infectious virus particle). The present method further allows the incorporation of bioconjugation handles (up to two bioconjugation handles per capsid protein) enabling precise attachment of 15 or 20 bioconjugation groups per capsid.

[0012] More specifically, the present invention encompasses a method to produce AAV vectors, in which any one of, or combination of, the minor capsid proteins VP1 and VP2, can be site-specifically chemically modified. Since the three capsid proteins are present at different stoichiometry (5, 5, and 50 copies of VP1, VP2 and VP3 respectively per capsid), this method provides the ability to create homogeneous AAV conjugates where modifications can be introduced into capsids with precise control over both the site and the number of modifications and retain biological activity of the virus, especially its targeted cellular infectivity.

[0013] Using engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that suppress a nonsense codon, as described herein, it is possible to site-specifically incorporate natural

and unnatural amino acids (UAAs) into the capsid proteins. Due to their small size, UAAs are well-tolerated in the AAV capsid at many positions.

[0014] Encompassed by the present invention is a genetically-modified adeno-associated virus (AAV) wherein the AAV capsid comprises at least one variant minor capsid protein VP1, VP2, or both VP1 and VP2, wherein the variant minor capsid protein is mutated at one, or more, amino acid residue sites to incorporate a natural amino acid not present in the wild-type AAV, or an unnatural amino acid (UAA) relative to the wild-type AAV VP1 or VP2 capsid protein. Importantly, the genetically-modified AAV s described herein retain their infectivity as compared to the wild-type AAV under comparable conditions.

[0015] In particular, the present invention encompasses a genetically-modified adeno-associated virus (AAV) comprising a variant minor capsid protein, wherein the capsid coding sequence (SEQ ID NO:1) is mutated at the translation origin of the AAV VP1 or VP2 or VP3 capsid protein open reading frame (ORF) to prevent translation of VP1, VP2 or both VP1 and VP2, resulting in an AAV capsid protein with VP1, VP2 , VP3 or both VP1 and VP2 deleted. The deleted/missing capsid protein is then provided (i.e, expressed) *in trans* from a second capsid coding sequence encoding the deleted minor capsid protein(s) resulting in a genetically-modified AAV with one, or more variant capsid proteins. VP2 is also represented by SEQ ID NO:2 and VP3 is also represented by SEQ ID NO:3 herein.

[0016] The genetically-modified AAV capsid protein comprises SEQ ID NO:1, or a sequence comprising at least about 70%, 75%, 80%, 85%, 90%, 95% , 99%, or between 90% and 99%, sequence identity of SEQ ID NO:1. The variant capsid protein is mutated at specific sites to delete VP1, VP2 or both VP1 and VP2 (e.g., as shown in Figure 2). For example, at position T454 of SEQ ID NO:1 a stop codon can be inserted using the techniques described herein. Such techniques are also described in, for example, PCT/US2020/055834; PCT/US2020/038766 or PCT/US2020/029567, the teachings of which are incorporated herein in their entirety, by reference. More specifically, the stop codon incorporated in the variant capsid protein can be a TAG, TAA or TGA suppressor stop codon.

[0017] Specifically encompassed by the present invention is the minor capsid protein CMV-VP1-delVP23 (FIG, 20, SEQ ID NO: 7)); CMV-VP2-delVP3 (FIG. 21, SEQ ID

NO:9) and CMV-VP1-VP2-delVP3 (FIG. 22, SEQ ID NO:10). As described herein, the CMV promoter sequence is illustrated but any suitable promoter sequence can be incorporated. Additionally, the minor capsid proteins can incorporate naturally-occurring or unnatural amino acids as described herein.

[0018] It is important to note that a key advantage of the genetically-modified AAV of the present invention is that the UAAs incorporated into the minor capsid proteins are at low levels, so their production efficiency is not compromised) significantly affected) by UAA incorporation. Thus, the AAV is packaged/assembled with infectivity titers substantially comparable to wild-type virus.

[0019] The variant capsid protein of the genetically-modified AAVs of the present invention can comprise one, or more mutated amino acid residues of VP1, VP2 or both VP1 and VP2 wherein the mutated amino acid residue site incorporates an unnatural amino acid (UAA). Examples of UAAs suitable for use in the present invention are shown in Figure 1, showing formulas 1-12 (upper line 1 formulas 1-6, left to right and second line formulas 7-12 left to right, including length of carbon chain extensions and substitutions. More specifically, unnatural amino acids suitable for use in the present invention can be selected from the group consisting of phenylalanine analogs; tyrosyl analogs; tryptophanyl analogs; or lysyl analogs. In particular the analogs are selected from the group consisting of: p-benzoylphenylalanine (pBpA); O-methyltyrosine (OMeY); 5-azidotryptophan; 5-propargyloxytryptopha; 5-aminotryptophan; 5-methoxytryptophan; 5-O-allyltryptophan; 5-bromotryptophan; azido-lysine (AzK) or Nε-acetyllysine (AcK)); C5Az; LCA; Nε-acetyllysine (AcK); cyclopropene amino acid, N^{ε}-(1-methylcycloprop-2-enecarboxamido)-lysine (CpK); 5-hydroxy-tryptophan (5-HTP);LCAlk; DiaazK and LCKet.

[0020] Alternatively, the mutation of the variant minor capsid protein of the genetically-modified AAV can be a natural amino acid residue, such as a cysteine or a selenocysteine. Specifically, the natural amino acid residue is different than the amino acid residue located at the same position of the wild-type AAV.

[0021] The genetically-modified AAV of the present invention can further comprise a bioconjugation handle. Such a bioconjugation handle can be covalently attached to the mutated natural or unnatural amino acid residue of the variant minor capsid protein in a site-specific manner. As described herein, 5 to 10 bioconjugation handles can be specifically incorporated into the minor capsid proteins, thus resulting in a genetically-

modified AAV with 5-10 bioconjugation handles per fully-assembled AAV capsid. Again, the genetically-modified AAV comprising at least one variant minor capsid protein VP1, VP2 or both VP1 and VP2, retains infectivity of target cells comparable to wild-type AAV under similar conditions, such as cell culture conditions. For example, the mutated amino acid can be conjugated with a chemical or biological entity (e.g., a biological entity can comprise a protein, peptide, nucleic acid, lipid or carbohydrate).

[0022] Specifically encompassed by the present invention, is a genetically-modified infectious AAV comprising a capsid comprising variant capsid proteins. For example, the variant capsid protein can comprise SEQ ID NO:1, or a sequence with at least about 80% sequence identity with SEQ ID NO:1, wherein the sequence is mutated to delete VP1, VP2 or both VP1 and VP2. Mutations of the capsid of the genetically-modified AAV can be at position(s), for example, 263, 454, 456, 587 and/or 588 wherein the mutated positions are relative to the wild-type VP1 sequence. The VP1 sequence can also be mutated at positions 263, 454, 456 and 588 but not 587. Alternatively, the VP1 sequence can be mutated at positions 263, 454, 456 but not 585 nor 588.Characteristics of the infectious genetically-modified AAV of the present invention are infectivity substantially comparable to the infectivity of wild-type AAV under comparable conditions, packaging of AAV with titers substantially comparable to wild-type AAV under comparable conditions, or both infectivity and packaging comparable to wild-type AAV.

[0023] In one embodiment of the present invention the mutated amino acid residue(s) of the modified AAV is functionalized or conjugated with a chemical or protein entity (also referred to herein as a bioconjugation handle or group). As described herein, the chemical or protein entity is selected from the group consisting of probes, small molecule ligands, peptides, cyclic peptides, nucleotides, polymers, proteins, or a virus conjugate. In a particular embodiment, the genetically-modified infectious AAV comprises a chemical or protein entity, wherein the entity is a cyclic peptide cRGD polyethylene glycol (PEG).

[0024] In another embodiment of the present invention, the mutated amino acid residue site of the variants VP1 or VP2 incorporates a naturally-occurring amino acid. In particular, the naturally-occurring amino acid is cysteine or selenocysteine.

[0025] The genetically-modified infectious AAV of the present invention comprising the variant capsid protein(s) (for example, the genetically-modified AAV comprising a mutated capsid protein functionalized with a cRGD peptide) can be "re-targeted" from

binding to, or recognizing or interacting with, its native cognate cellular receptor to bind to, or recognize or interact with a non-cognate targeted cell. As described herein any number of retargeting groups attached per capsid dramatically affects the properties of the resulting AAV conjugates. This underscores the importance of controlling the site and the stoichiometry for altering the properties of AAV vectors by chemical modification. As described herein, it is now possible for precise, selective chemical modification of AAV capsids with unprecedented control over both the site and the stoichiometry (how many modifications are introduced per capsid). Such control will be extremely important for the development of engineered AAV vectors with more therapeutically desirable properties.

[0026] Also encompassed by the present invention is a method of producing/making an infectious genetically-modified AAV, wherein the AAV comprises a variant AAV capsid protein, wherein VP1, VP2 or both VP1 and VP2 comprise one, or more mutated amino acid residue sites, relative to wild-type VP1, VP2 or both VP1 and VP2. In general, the method comprises providing competent host cells in culture, transfecting the cultured cells with one, or more, plasmids comprising 1) an AAV variant VP3 that does not express VP1 or VP2 or neither VP1 and VP2; 2) a variant VP1, VP2 or both VP1 and VP2 with a suitable promoter that do not express VP3; 3) additional factors required for AAV expression ; 4) providing the required unnatural amino acid and 5) providing a plasmid encoding an engineered aminoacyl-tRNA synthetase/tRNA pair that selectively charge the unnatural amino acid in response to a stop codon. The cells and plasmids are cultured under conditions sufficient for expression of the plasmid genes and assembly of the AAV, thereby producing an infectious, genetically-modified AAV comprising a variant capsid protein, wherein VP1, VP2 or both VP1 and VP2 are mutated at one, or more amino acid residues relative to the wild-type AAV.

[0027] More specifically, the method comprises the following steps of providing competent host cells in culture; and co-transfecting the cultured cells with plasmids/constructs comprising the required sequences for assembly and expression of the genetically-modified infectious AAV. One mRNA encodes VP1, VP2 or VP3 with different start codon sites (see Figure 2). If a start site is mutated then expression of VP1, VP2 or VP3 can be controlled by expressing separate mRNAs for each capsid protein, and the capsid proteins can be selectively mutated yet the mutated minor capsid proteins VP1

and/or VP2 and the wild-type can be assembled resulting in an infectious genetically-modified AAV.

[0028] For example, as described herein, an unnatural amino acid can be incorporated site-specifically into VP1 and/or VP2 by co-transfecting competent host cells with plasmids encoding an engineered tRNA synthetase and tRNA pair that charge the unnatural amino acid. More specifically, provided is an orthogonal tRNA/aaRS pair comprising an engineered amino-acyl RNA synthetase (aaRS) and its corresponding tRNA to co-translationally incorporate an unnatural amino acid (UAA) in response to the mutated site of the variant VP1, VP2, VP3 or both VP1 and VP2, and providing the required unnatural amino acid to be incorporated at the mutated site.

[0029] Also provided (co-cultured) along with plasmids encoding the mutated VP1 and/or VP2 and wild-type VP3 transcripts are plasmid(s) encoding additional factors required for AAV capsid expression and assembly of the complete capsid (i.e., comprising VP1, VP2 and VP3) in cell culture. The cells, plasmids and UAAs are cultured/maintained under conditions sufficient for expression of the plasmid genes and assembly of the AAV, thereby producing an infectious genetically-modified AAV comprising a variant AAV capsid protein, wherein VP1, VP2 or both VP1 and VP2 are mutated at one, or more amino acid residue sites relative to the wild-type AVV.

[0030] The engineered aminoacyl-tRNA-synthetase-tRNA pair can be derived from the E.coli leucyl pair, E. coli tryptophanyl pair, E. coli tyrosyl pair or archea-derived pyrrolysyl pair.

[0031] The infectious, genetically-modified AAV can be recovered/harvested from the cell culture and evaluated for biological activity using techniques as described herein and as known to those of skill in the art.

[0032] The mutated amino acid residue site of the infectious genetically-modified AAV can incorporate an unnatural amino acid (UAA) analog as shown in FiG. 1, formulas 1-12. More specifically, the unnatural amino acid is selected from the group consisting of phenylalanine analogs; tyrosyl analogs; tryptophanyl analogs; or lysyl analogs. The analogs are selected from the group consisting of: p-benzoylphenylalanine (pBpA); O-methyltyrosine (OMeY); 5-azidotryptophan; 5-propargyloxytryptopha; 5-aminotryptophan; 5-methoxytryptophan; 5-O-allyltryptophan; 5-bromotryptophan; azido-lysine (AzK) or Nε-acetyllysine (AcK); C5Az; LCA; Nε-acetyllysine (AcK); cyclopropene amino acid, N^{ε}-(1-

methylcycloprop-2-enecarboxamido)-lysine (CpK); 5-hydroxy-tryptophan (5-HTP);LCAlk; DiaazK and LCKet.

[0033] Alternatively, the mutated amino acid residue site of the variant capsid protein incorporates a naturally-occurring amino acid, and the naturally-occurring amino acid is cysteine or selenocysteine.

[0034] The mutated amino acid of the variant capsid protein can be functionalized with a chemical or protein entity, wherein the entity is selected from the group consisting of probes, small molecule ligands, peptides, cyclic peptides, nucleotides, polymers, proteins, or a virus conjugate. In one embodiment the entity functionalizing the mutated amino acid of the variant capsid protein of the infectious genetically-modified AAV is a cyclic peptide cRGD.

[0035] The infectious genetically-modified AAV can comprise the mutated VP1 amino acid sequence comprises SEQ ID NO:1, or a sequence with at least about 80% sequence identity with SEQ ID NO:1, wherein the variant VP1 capsid protein is mutated at one, or more position(s) located at 263, 454, 456, 587 and/or 588 of the VP1 sequence and relative to the wild-type VP1 sequence. The VP1 sequence can also be mutated at positions 263, 454, 456 and 588 but not 587. Alternatively, the VP1 sequence can be mutated at positions 263, 454, 456 but not 585 nor 588.

[0036] The infectious genetically-modified AAV can comprise the mutated VP2 amino acid sequence comprises SEQ ID NO:2, or a sequence with at least about 80%, 85%, 90%, 95 or 99% sequence identity with SEQ ID NO:2.

[0037] The infectious genetically-modified AAV can comprise the VP3 amino acid sequence comprises SEQ ID NO:3, or a sequence with at least about 80%, 84%, 90%, 95% or 99% sequence identity with SEQ ID NO:3.

[0038] Also encompassed by the present invention are therapeutic or antigenic compositions comprising a genetically-modified AAV as described herein and further comprising one, or more therapeutic or antigenic gene constructs suitable for gene therapy or vaccination (e.g., eliciting an immune response in a subject). If the composition is a vaccine composition comprising an antigen, the composition can further comprise an adjuvant.

[0039] Methods of treating a disease or condition in a subject are also encompassed herein. For example, the method of treatment can comprise administering a therapeutic composition to the subject, wherein the composition comprises an AAV vector as described herein and a gene construct encoding a protein or peptide in a therapeutic amount capable of decreasing or alleviating the disease or condition in the subject. Use of a genetically-modified AAV vector of the present invention in therapeutic compositions can be particularly useful for treating cancer, or a disease such as cystic fibrosis of sickle-cell anemia, where providing a replacement gene encoding a functional protein in a targeted manner would decrease, or completely alleviate, the cancer or disease symptoms.

[0040] Another example is a method of eliciting an immune response in a subject, the method comprising administering an antigenic/vaccine composition to the subject, wherein the antigenic composition comprises the genetically-modified AAV vector of the present invention and a gene construct encoding an antigenic protein or peptide capable of eliciting an immune response in the subject. Such a method of using the AAV vector of the present invention would specifically target immune cells capable of mounting an immune response.

[0041] Kits comprising the genetically-modified AAV of the present invention are also encompassed herein. Such kits can include a suitable vial containing the genetically-modified AAV as described herein, for example a vial of the AAV in a sterile diluent, as well as vials/containers of supplemental components, for example for propagating the AAV in cell culture. An instructional pamphlet can also be included in the kit.

[0042] The above and other features of the invention including various novel details of construction and combinations of parts, and other advantages, will now be more particularly described with reference to the accompanying drawings and pointed out in the claims. It will be understood that the particular method and device embodying the invention are shown by way of illustration and not as a limitation of the invention. The principles and features of this invention may be employed in various and numerous embodiments without departing from the scope of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] In the accompanying drawings, reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale; emphasis has

instead been placed upon illustrating the principles of the invention. The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings(s) will be provided by the Office upon request and payment of the necessary fee. Of the drawings:

[0044] Figure 1 shows the examples of the structures of the natural and the unnatural amino acids that can be incorporated into the capsid of AAV in a controlled stoichiometry. Specifically shown are UAA analogs comprising the formulas 1-12 (upper line 1 formulas 1-6, left to right and second line formulas 7-12 left to right, including length of carbon chain extensions and substitutions.

[0045] Figure 2 shows the scheme for introducing mutations (natural or unnatural amino acids) selectively into VP1, or VP2, or VP1 and VP2. Three capsid proteins are expressed from the same open reading frame Cap via the use of alternative splicing and start codon usage. Mutations have been engineered at the translation origin (demonstrated by red crosses) that prevent the expression of VP1, VP2, or VP3 from this ORF. The missing minor capsid protein(s) can then be supplied *in trans*, driven by a strong promoter such as CMV. Separating the expression of VP1, VP2, or VP1+VP2 from the rest of the capsid proteins makes it possible to selectively mutate these without affecting the other capsid proteins.

[0046] Figure 3 shows packaging of AAV2 in HEK293T cells using constructs described in Figure 2. AAV with variant capsid proteins have comparable yields relative to the original system.

[0047] Figure 4 shows the infectivity of the packaged, tittered viruses produced in Figure 3 at constant titer measured by their ability to deliver and express an EGFP reporter gene in HEK293T cells.

[0048] Figure 5 shows selective unnatural amino acid mutagenesis of individual minor capsid proteins in AAV capsid and their use to chemically attach a fluorophore.

[0049] Figure 6 shows the number of retargeting ligands attached to the AAV capsid dramatically affects its retargeting efficiency.

[0050] Figure 7A-7C show precise labeling of AAV at engineered cysteine residues.

[0051] Figure 8A-C show the amino acid sequences for AAV isoform VP1(SEQ ID NO:1). Figures 8B and 8C disclose SEQ ID NO: 2 and SEQ ID NO: 3, respectively.

[0052] Figure 9 shows the amino acid sequence of AAV capsid protein isoform VP2. (SEQ ID NO:2).

[0053] Figure 10 shows the amino acid sequence of the AAV capsid protein isoform VP3. (SEQ ID NO:3).

[0054] Figure 11A-C shows the selective incorporation of the UAA C5Az either VP1 or VP2 or VP1+VP2Fig.

[0055] Figure 12A-C shows the results of LCA incorporated into VP1.

[0056] Figure 13A-C shows the results of incorporation of CpK into VP1.

[0057] Figure 14A-C shows the results of incorporation of 5HTP into VP1.

[0058] Fig. 15A-B demonstrates incorporation of several other unnatural amino acids, LCAlk, DiazK and LCKet, into VP1.

[0059] Fig. 16A-B shows the results of selective PEGylation of AAV at VPI site 454.

[0060] Figure 17 shows the nucleic acid sequence encoding RC2-VP1-del. (SEQ ID NO:4) The location of the mutation is shown in red.

[0061] Figure 18 shows the nucleic acid sequence encoding RC2-VP2-del. (SEQ ID NO:5) The location of the mutation is shown in red.

[0062] Figure 19 shows the nucleic acid sequence encoding RC2-VP12-del. (SEQ ID NOL6) The location of the mutation is shown in red.

[0063] Figure 20 shows the nucleic acid sequence encoding CMV-VP1-delVP23. (SEQ ID NO:7) The locations of the mutations are shown in red.

[0064] Figure 21 shows the nucleic acid sequence encoding CMV-VP2-delVP3. (SEQ ID NO:8) The locations of the mutations are shown in red.

[0065] Figure 22 shows the nucleic acid sequence of CMV-VP1-VP2-delVP3. (SEQ ID NO:9) The locations of the mutations are shown in red.

[0066] Figure 23A is the plasmid map of pIDTsmart-ITR-GFP-4xEcLtR-LeuRS. Figure 23B shows the nucleic acid sequence of the plasmid. (SEQ ID NO:10).

[0067] Figure 24A shows the plasmid map of pIDTsmart-TrpRS-8xWtR-ITR-GFP. Figure 24B shows the nucleic acid sequence of the plasmid. (SEQ ID NO:11).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0068] The invention now will be described more fully hereinafter with reference to the accompanying drawings, in which illustrative embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art.

[0069] As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items. Further, the singular forms and the articles "a", "an" and "the" are intended to include the plural forms as well, unless expressly stated otherwise. It will be further understood that the terms: includes, comprises, including and/or comprising, when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. Further, it will be understood that when an element, including component or subsystem, is referred to and/or shown as being connected or coupled to another element, it can be directly connected or coupled to the other element or intervening elements may be present.

[0070] It will be understood that although terms such as “first” and “second” are used herein to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another element. Thus, an element discussed below could be termed a second element, and similarly, a second element may be termed a first element without departing from the teachings of the present invention.

[0071] Unless otherwise defined, all terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and will not be interpreted in an idealized or overly formal sense unless expressly so defined herein.

[0072] It should be noted that this invention description is not limited to these specific methodologies, compositions, cell lines/biological systems, or any other standard

protocol. The present invention discloses a general platform for producing AAV vectors that can be chemically functionalized using chemo-selective reactions with control over site and stoichiometry of modification, the cell type used for virus packaging, identity of the engineered aminoacyl-tRNA synthetase (aaRS) or tRNA, or UAAs or natural amino acids, the chemical reaction used for introducing the modification, etc., can vary as the technology in this field and in this work advance.

[0073] Previously, incorporation of UAAs into the AAV capsid has been directed to all 60 capsid proteins. Since the three overlapping capsid proteins are expressed from the same open reading frame (ORF) Cap via alternative splicing and start codon usage, it has not been possible to selectively modify a subset of these proteins. However, subsequent chemical modification of all of the 60 capsid proteins perturbs the infectivity of the virus (e.g., Figure 6, blue trace). While the mechanism of this perturbation is poorly understood, it is reasonable to believe that the molecular processes associated with the complex entry pathway of the virus can be affected by over-modification of the capsid. Additionally, it is desirable to be able to create AAV vectors into which a defined number of chemical modifications can be introduced in a site-specific manner. Indeed, retargeting experiments demonstrate that there is an optimal number of ligands per capsid needed for efficient retargeting (Figure 6). It is possible to control the degree of modification of an AAV capsid with 60 UAA-handles by controlling the concentration of the modifying reagent or reaction time (Figure 6). However, this leads to a heterogeneous mixture of capsids in different states of modification. The ability to create homogeneous AAV conjugates where a defined number of modifications are introduced per capsid is critical for gene therapy.

[0074] In the AAV capsid, the three capsid proteins VP1, VP2, and VP3 are incorporated at a roughly 1:1:10 stoichiometry. Consequently, approximately 5 copies of each of the two minor capsid proteins, VP1 and VP2, are present in the capsid, with 10 copies of VP3, for a total of 60 copies of capsid proteins per virus particle. The ability to selectively introduce engineered, modifiable natural or unnatural amino acid residues into these minor capsid proteins provides an avenue to introduce a controlled number of handles per capsid. However, since the AAV capsid genes are all encoded in the same ORF and expressed by alternative splicing and start-codon usage, it is challenging to mutate one capsid protein without affecting the others. The present invention describes a method to separately express the minor capsid proteins (VP1, or VP2, or VP1+VP2) by introducing

mutations at the translation origins of VP1 and/or VP2, such that these are not expressed from the native Cap ORF. Next, the missing capsid protein can be expressed *in trans* from a strong promoter, for example, the CMV promoter as described herein. Expression of the undesired capsid proteins (e.g., VP3) from this second ORF is also similarly eliminated by mutating translation origins (Figure 2). This platform provides the ability to express any combinations of the three capsid proteins from one ORF and separating the expression of a chosen third from a second ORF, making it possible to selectively engineer any subset of the three capsid proteins.

[0075] Using the platform described herein, it is now possible to selectively introduce unique non-coding codons (such as nonsense, 4-base, or unnatural base-pair-containing codons) into any combination of the three capsid proteins by site-directed mutagenesis. When such a mutant Cap is co-expressed with a suitable engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pair that can decode the unique codon, an UAA residue can be incorporated into these sites.

[0076] In one embodiment of this invention, the UAA AzK was incorporated at residue T454 (VP1 numbering) of VP1 or VP2 or VP1+VP2 (Figure 3). Using a pyrrolysyl-tRNA synthetase/tRNA pair. The resulting viruses were packaged at an efficiency comparable to the wild-type virus and had comparable infectivity, demonstrating these modifications are well-tolerated (Figure 3 and Figure 4). The importance of controlling the number of modifications per capsid was demonstrated by the differential behavior of AAV where the UAA was incorporated at 5 copies per capsid (mutation of either VP1 or VP2 alone), or 10 copies per capsid (mutation of both VP1+VP2), or 60 copies per capsid (all three capsid proteins were mutated), and a retargeting ligand was subsequently attached to these UAA handles (Figure 6). AAV with 60 attachment handles enabled efficient retargeting at intermediate modification states, but upon full modification, all infectivity was lost. 5 or 10 modifications per capsid were well-tolerated, but 10 retargeting ligands per capsid was needed for efficient retargeting (Figure 6).

[0077] It should be noted that any other combination of capsid proteins, and any site within these proteins, can be engineered using this strategy. Especially well-suited for mutation is the N-terminus of the capsid proteins as this section of the capsid protein is essentially internal when folded into the functional capsid, whereas the C-terminus of the capsid protein is exposed.

[0078] Additionally, other aaRS/tRNA pairs can be used to (including, but not limited to, bacterial tyrosyl, tryptophanyl, and leucyl-tRNA synthetase/tRNA pairs) and incorporate any other unnatural amino acids (illustrative examples shown in Figure 1 and described herein). This technology can also be extended to any other natural serotype of AAV as well as engineered and evolved variants of AAV.

[0079] Cysteine and selenocysteine are natural amino acid residues found in proteins. Because of their low abundance and unique reactivity, these can be used for site-selective bioconjugation reactions. In another embodiment of this invention, engineered surface-exposed cysteine residues can be introduced into the minor capsid proteins (Figure 7A and 7B). Even though the same mutation is not well-tolerated when introduced to all 60 capsid proteins, leading to low titer and poor infectivity, robust virus packaging and infectivity was observed when surface exposed cysteine residue was introduced only to VP1 or VP2 at the T454 site. The ability to selectively modify the engineered cysteine residue on the minor capsid protein was also demonstrated (Figure 7C).

[0080] Because of the complexities associated with the assembly and the cell-entry process, for which the AAV capsid has been optimized through evolution, it frequently resists attempts at engineering the capsid protein through natural/unnatural amino acid mutagenesis, loop insertion, protein fusion etc. However, engineering just the minor capsid proteins is tolerated significantly better as it introduces much less overall perturbation to the capsid overall. This is illustrated by the tolerance of the engineered cysteine residue at the minor capsid protein, but not everywhere (Figure 7). Thus, this invention provides an opportunity to introduce more aggressive engineering to alter the properties of the AAV capsid through engineering the minor capsid protein. In addition to natural and unnatural amino acid mutagenesis, this invention can also be used to introduce peptide and protein fusions and insertions into the minor capsid proteins that either directly provide a beneficial trait (e.g., binding a certain target), or can be selectively modified through chemical or enzymatic reactions (e.g., biotinylation tag, SNAP or HALO tag, etc.).

[0081] This invention allows the introduction of a defined number of engineered residues (natural or unnatural) per capsid by selectively mutagenizing VP1, or VP2, or VP1 plus VP2. Further control over the number of engineered sites can be achieved by introducing more than one engineered residue into VP1, or VP2, or VP1+VP2. As described herein, various sites of the VP1 capsid protein have been selectively mutated

such as 263, 454, 456, 587 and 588 (the numbering corresponds to the amino acid residues of wild-type VPI). The platform can be extended to any packaging platform including, but not limited to, mammalian cells, insect cells, and cell-free translation/packaging systems.

[0082] The invention enables the incorporation of numerous natural and unnatural amino acid residues into AAV with control over site and copy number, with a wide variety of different chemistries which can be used to chemo-selectively attach various entities. In one embodiment of this invention, an azido-containing UAA was introduced selectively into the minor capsid proteins of AAV, followed by their conjugation to a fluorophore or a retargeting ligand using strain-promoted azide-alkyne click reaction (Figure 5 and Figure 6). In another embodiment, an engineered cysteine residue in a minor capsid protein was introduced and subsequently conjugated to a fluorophore using cysteine-maleimide coupling reaction. Any other chemo-selective conjugation reaction can be applied for the capsid modification including, but not limited to, inverse-electron demand Diels-Alder reaction between a strained alkene and a tetrazine, or a furan and a maleimide, condensation reaction between an aldehyde/ketone and an amino-oxy/hydrazine groups, chemo-selective rapid azo-coupling reaction (CRACR), oxidative and photocatalyzed coupling reactions, nucleophilic substitution/addition by cysteine or selenocysteine residue to various electrophiles, etc.

[0083] The methods described herein can be used to attach a wide variety of entities including, but not limited to, probes (fluorescent, radioactive, MRI, luminescent, etc.), small molecule ligands, peptides, cyclic peptides, nucleotides (DNA, RNA, LNA, PNA, etc.), polymers (such as PEG), carbohydrates (e.g., sialic acids, etc.), proteins (e.g., enzymes, nanobodies, antibodies, etc.), another AAV of the same or different serotype, etc. Such attachment can provide AAV conjugates that efficiently retargets to a user-defined receptor, thus altering its tissue tropism. Immune-evading AAV can also be created by site specifically attaching groups (such as PEG, peptides, carbohydrates, or other polymers) that passively protect the capsid from the immune system, or ligands that actively bind inhibitory receptors on immune cells to turn off immune response (such as SIGLEC ligands). It can also be used to attach enzymes on AAV capsids to produce capsids with superior infectivity profile, or conjugate two AAVs with same or different serotypes to create novel class of vectors with expanded cargo capacity as well as novel tropism.

[0084] The controlled AAV modification technology descried herein can be used for many applications such as targeting AAV vectors to desired types of cells by attaching retargeting ligands that include, but are not limited to, small molecules, peptides, cyclic peptides, nanobodies, antibodies and antibody fragments; DNA/RNA/PNA aptamers, etc.; optimizing the properties of such conjugates by systematically controlling the attachment site, the number of retargeting groups per capsid, and the chemical properties of the linker; Attenuating the immune response generated by AAV vectors by the controlled attachment of immuno-modulatory entities including, but not limited to, polyethylene glycol and other polymers, carbohydrates (such as sialic acid, etc.), ligands that bind inhibitory receptors on immune cells (such as SIGLEC receptor), etc.

[0085] In particular the genetically-modified AAV of the present invention can be used as an enhance vector for gene therapy, Using techniques described herein, the AAV can be genetically-modified to specifically target/direct delivery of a therapeutic or antigenic gene construct to the target cell with enhanced/increased efficacy over non-modified AAV vectors. The cell can be cultured *in vitro* or *in vivo* or *ex vivo* delivery to a subject in need thereof. The term "subject" as used herein can include any animal subject, and in particular includes a mammalian subject such as a human. The human subject can be treated for medical purposes using the AAV gene vector described herein, such as for the treatment of an existing disease, disorder, condition or the prophylactic treatment for preventing the onset of a disease, disorder, or condition or an animal subject for medical, veterinary purposes, or developmental purposes. The methods and compositions of the present invention may be used to treat any type of cancerous tumor or cancer cells. Additionally, the genetically-modified AAV of the present invention can be used in a vaccine composition wherein a nucleic acid sequence encoding an antigenic agent such as a protein or peptide is delivered to the targeted cell along with additional components such as adjuvants wherein an immune response is elicited in the subject.

[0086] The gene construct delivered by the genetically- modified AAV is therapeutically effective. A “therapeutically effective” amount as used herein refers to an amount sufficient to have the desired biological effect to produce the desired effect on the underlying disease state (for example, an amount sufficient to inhibit tumor growth in a subject, produce an immune response to an antigen or to inhibit autoimmune disease) in at least a sub-population of cells in a subject at a reasonable benefit/risk ratio applicable to

any medical treatment. Determination of therapeutically effective amounts of the constructs/agents used in this invention, can be readily made by one skilled in the art, by the use of known techniques and by observing results obtained under analogous circumstances.

[0087] The technology can also be used for conjugating two distinct AAV vectors, of the same or different serotypes, using a bifunctional linker to increase the overall size of the genetic cargo delivered per cell. Such novel conjugates between two different AAV vectors also will have unique tropism, and for conjugating external payloads (such as proteins, small molecules, nucleic acids, probes, etc.) onto the virus capsid to be delivered into cells in vitro and in vivo for research or therapeutic purposes that work independently, or in conjunction with the genetic cargo inside the AAV capsid.

[0088] Finally, this technology can be used for the investigation of the entry pathway of AAV capsids into mammalian cells by incorporating groups such as fluorescent probes, photo-crosslinkers, and affinity handles (such as biotin).

[0089] Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present invention to its fullest extent. The following specific embodiments and examples are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

[0090] Incorporation of UAAs into the capsid of AAV requires efficient expression of its capsid protein(s), including the desired UAA modification in a competent cell (e.g., mammalian cell) serving as the host for virus amplification. As described earlier, the site of UAA incorporation can be specified by a stop codon (such as TAG), and the UAA of interest can be delivered by an engineered tRNA/aminoacyl-tRNA synthetase pair, with a cognate anticodon. Consequently, the production of UAA-modified viruses must involve the simultaneous expression of genetic components necessary for virus amplification, as well as those necessary for the amplification of the virus.

[0091] To incorporate an UAA into the capsid proteins of AAV, the replacement of several surface exposed endogenous amino acids residues was targeted on the AAV capsid (guided by a crystal structure of the virus particle, see, e.g., Xie, et al., The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy" PNAS August 6, 2002: vol. 99: no. 16, p. 10407). Since such capsid mutations can potentially perturb viral infectivity, regions that are known to tolerate alterations were targeted, such

as those on the threefold proximal spike. The replacement of essential arginine residues was also targeted in the heparan sulfate receptor binding region, which disrupts native tropism and will facilitate re-targeting. This allows "erasing" the native host cell preference of AAV and rewriting it with targeting agents with precise labeling methodology. For example, positions targeted can reside in the domain conserved among all of the three AAV capsid proteins, leading to their replacement with UAA in all of them. AAV can be produced by transfecting e.g., HEK293T cells with plasmids containing the required elements in the presence of the UAA azido-lysine ("AzK"). In a specific example a Methanosarcina barkeri-derive pyrrolysyl-tRNA synthetase and an M. mazeii derived pyrrolysyl tRNA (TAG suppressor) was used to incorporate the AzK amino acid in response to the stop codon, TAG at T454.

[0092] The plasmid maps and sequences used in the examples described below are found in Figures 17-24.

[0093] **Example 1: Packaging of AAV2**

[0094] Figure 3 shows packaging of AAV2 in HEK293T cells using constructs described in Figure 2 to comparable yields relative to the original system. The qPCR titer of the original (WT) AAV2 is shown and the titer for the rest are shown relative to the WT. ΔVP1 and ΔVP2 represents AAV2 packaged using Cap genes from which the expression of VP1 and VP2, respectively, were eliminated. The virus still assembles efficiently in the absence of the minor capsid proteins. ΔVP1-CMV-VP1 and ΔVP2-CMV-VP2 represents AAV2 packaged using Cap genes from which the expression of VP1 and VP2, respectively, were eliminated and the respective proteins were expressed *in trans* from a CMV promoter. In these systems, the T454 residue in VP1 or VP2 were mutated to TAG and suppressed using a pyrrolysyl-tRNA synthetase/tRNA pair to incorporate an UAA (AzK), which are represented by ΔVP1-CMV-VP1-454AzK and ΔVP2-CMV-VP2-454AzK, respectively. Packaging yield of the virus in the presence or the absence of the UAA added to the media is shown.

[0095] **Example 2: Infectivity of the Packaged Virus**

[0096] Figure 4 shows the infectivity of the packaged, tittered viruses produced in Figure 3 at constant titer measured by their ability to deliver and express an EGFP reporter gene in HEK293T cells. While ΔVP1 and ΔVP2 viruses package well, these show significantly attenuated infectivity. Supplying VP1 and VP2 in trans (ΔVP1-CMV-VP1

and ΔVP2-CMV-VP2) rescues the infectivity. ΔVP1-CMV-VP1-454AzK and ΔVP2-CMV-VP2-454AzK viruses produced in the presence of the UAA show robust infectivity, while those in the absence does not. It should be noted that in the absence of UAA, the TAG mutants of the minor capsid proteins will fail to express.

[0097] **Example 3: Selective Fluorophore Labeling of Mutated AAV**

[0098] Figure 5 shows selective unnatural amino acid mutagenesis of individual minor capsid proteins in AAV capsid and their use to chemically attach a fluorophore. Following purification, different recombinant AAV2 preparations were labeled with a cyclooctyne-fluorophore, which selectively labels the azide group present in the UAA AzK. The top panel shows SDS-PAGE analysis of the AAV2 preparations, the bottom panel shows the fluorescence image of the same gel. As expected, the wild-type AAV2 does not show labeling, T454AzK (AzK in all 60 capsid proteins) show labeling of all the capsid proteins, while ΔVP1-CMV-VP1-454AzK and ΔVP2-CMV-VP2-454AzK show selective labeling of VP1 and VP2, respectively, thus demonstrating our ability to selectively label distinct minor capsid proteins.

[0099] **Example 4: Retargeting Efficiency of Mutated AAV**

[00100] Figure 6 shows the number of retargeting ligands attached to the AAV capsid dramatically affects its retargeting efficiency. Attaching cRGD ligands onto detargeted AAV2 capsids (where binding of the natural primary receptor, heparin sulfate proteoglycan receptor HSPG, was ablated by mutating key residues R588 and R587 to Ala; designated by amino acid residue location of VP1) enables it to selectively bind and infect cancer cell-lines such as SK-OV-3 that overexpress the αVβ3 integrin receptor. These graphs show the infectivity of the shown detargeted AAV2 mutants as they were incubated with cyclooctyne-cRGD over time, as it progressively functionalizes the AzK side chains with cRGD. For 454AA (60 AzK per capsid), the infectivity toward SK-OV-3 cells first go up and then come back down, suggesting an optimal number of cRGD per capsid needed for efficient retargeting. Over-modification of the capsid at later times leads to a loss of infectivity, likely by perturbing viral entry processes. For VP1-454AA and VP2-454AA (5 AzK each per capsid), infectivity goes up and reaches a plateau upon prolonged incubation, suggesting that attachment of 5 cRGD per capsid is not detrimental to AAV2 infectivity. However, the maximal infectivity reached for these mutants are low, suggesting that 5 cRGDs per capsid may not be sufficient for efficient retargeting. VP1+2-454AA (10 AzK

per capsid) behaved just like VP1-454AA and VP2-454AA, but the infectivity upon prolonged incubation reaches levels similar to the optimal infectivity observed with T454-AA at the optimal degree of modification.

[00101] **Example 5: Precise Labeling of Engineered Cysteine Residues of AAV**

[00102] Figure 7A-7C show precise labeling of AAV at engineered cysteine residues. (A) Packaging efficiency (qPCR) of AAV produced using wild-type Cap (WT), T454C mutant of Cap (T454C at all three capsid proteins), VP1-T454C (T454C mutant of the trans-substituted VP1), VP2-T454C (T454C mutant of the trans-substituted VP2). (B) Infectivity of these viruses measured by their ability to deliver and express an EGFP gene in HEK293T cells (FACS titer). (C) Selective labeling of the engineered 454-cysteine residue on VP1 by fluorescein-maleimide on VP1-454C virus. FAM-fluorescence image shows the result of the labeling reaction on WT and VP1-454C virus, whereas SYPRO stains all proteins.

[00103] **Example 6: Selective Incorporation of the UAA Into Either VP1 or VP2 or VP1+VP2**

[00104] As shown in Figure 11A-C, C5Az was incorporated into either VP1 or VP2 or VP1+VP2 by co-transfecting the HEK293T cells with a plasmid (pIDTsmart-ITR-GFP-4xEcLtR-LeuRS) encoding an engineered E. coli leucyl-tRNA synthetase (EcLeuRS) and tRNA pair that charge C5Az, as well as plasmids encoding [RC2-VP1-del + CMV-VP1-delVP23] or [RC2-VP2-del + CMV-VP2-delVP3] or [RC2-VP12-del + CMV-VP1-VP2-delVP3], respectively. The 454 position (VP1 numbering) in the desired protein was replaced with a TAG stop codon. A) Structure of C5Az, B) qPCR titers of the various preparations shown relative to wild-type AAV2 production, in the presence or absence of C5Az. C) Selective fluorescent labeling of VP1, using DBCO-rhodamine, on AAV2-VP1-454-C5Az.

[00105] **Example 7: Selective Incorporation of CpK Into VP1**

[00106] As shown in Figure 13A-C, CpK was incorporated into VP1 by co-transfecting the HEK293T cells with a plasmid (pIDTsmart-ITR-GFP-4xEcLtR-LeuRS) encoding an engineered E. coli leucyl-tRNA synthetase (EcLeuRS) and tRNA pair that charge CpK, as well as plasmids encoding [RC2-VP1-del + CMV-VP1-delVP23]. Various sites (as indicated; VP1 numbering) in VP1 was replaced with a TAG stop codon. A) Structure of

CpK, B) Relative titers of the various preparations shown relative to wild-type AAV2 production, in the presence or absence of LCA. C) Selective fluorescent labeling of VP1, using tetrazine-FITC, on AAV preparations incorporating LCA into site 456 of VP1.

[00107] **Example 8: Selective Incorporation of 5-HTP into VP1**

[00108] As shown in Figure 14A-C, 5HTP was incorporated into VP1 by co-transfecting the HEK293T cells with a plasmid (pIDTsmart-TrpRS-8xWtR-ITR-GFP) encoding an engineered E. coli tryptophanyl-tRNA synthetase (EcTrpRS) and tRNA pair that charge 5HTP, as well as plasmids encoding [RC2-VP1-del + CMV-VP1-delVP23]. Various sites (as indicated; VP1 numbering) in VP1 was replaced with a TGA stop codon. A) Structure of 5HTP, B) Relative titers of the various preparations shown relative to wild-type AAV2 production, in the presence or absence of 5HTP. C) Selective fluorescent labeling of VP1, using fluorescein amine and ferricyanide, on AAV preparations incorporating 5HTP into site 454 of VP1.

[00109] **Example 9: Incorporation UAAs into VPI**

[00110] As shown in Figure 15A-B, sseveral other unnatural amino acids were incorporated into VP1 by co-transfecting the HEK293T cells with a plasmid (pIDTsmart-ITR-GFP-4xEcLtR-LeuRS) encoding an engineered E. coli leucyl-tRNA synthetase (EcLeuRS) and tRNA pair that charge these unnatural amino acids, as well as plasmids encoding [RC2-VP1-del + CMV-VP1-delVP23]. Site 454 (VP1 numbering) in VP1 was replaced with a TAG stop codon. A) Structure of the unnatural amino acids, B) Relative titers of the various mutant AAV2 preparations shown relative to wild-type AAV2 production, in the presence or absence of the indicated unnatural amino acids.

[00111] **Example 10: Modification of the LCA Site of AAV2-VP1-454-LCA**

[00112] As shown in Figure 16A-B, AAV2-VP1-454-LCA, was selectively modified at the LCA site with 20 kDa polyethylene glycol (PEG) polymer using the corresponding PEG-DBCO conjugate. SDS-PAGE analysis shows selective labeling of VP1. Wild-type AAV2, AAV2-VP1-454-LCA, and PEG-modified AAV2-VP1-454-LCA show similar infectivity. Equal genome copies of each virus were added too HEK293 cells and the expression of the encoded luciferase reporter was monitored by a standard luciferase assay.

[00113] While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that

various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1. A genetically-modified adeno-associated virus (AAV) wherein the AAV capsid comprises at least one variant minor capsid protein VP1, VP2, or both VP1 and VP2, wherein the variant minor capsid protein is mutated at one, or more, amino acid residue sites to incorporate a natural amino acid or an unnatural amino acid (UAA) relative to the wild-type AAV VP1 or VP2 capsid protein.

2. The genetically-modified adeno-associated virus (AAV) of claim 1, comprising a variant minor capsid protein, wherein the variant capsid coding sequence is mutated at the translation origin of the AAV VP1, or VP2, or VP3 capsid protein open reading frames (ORF) to prevent translation of VP1, VP2 or both VP1 and VP2.

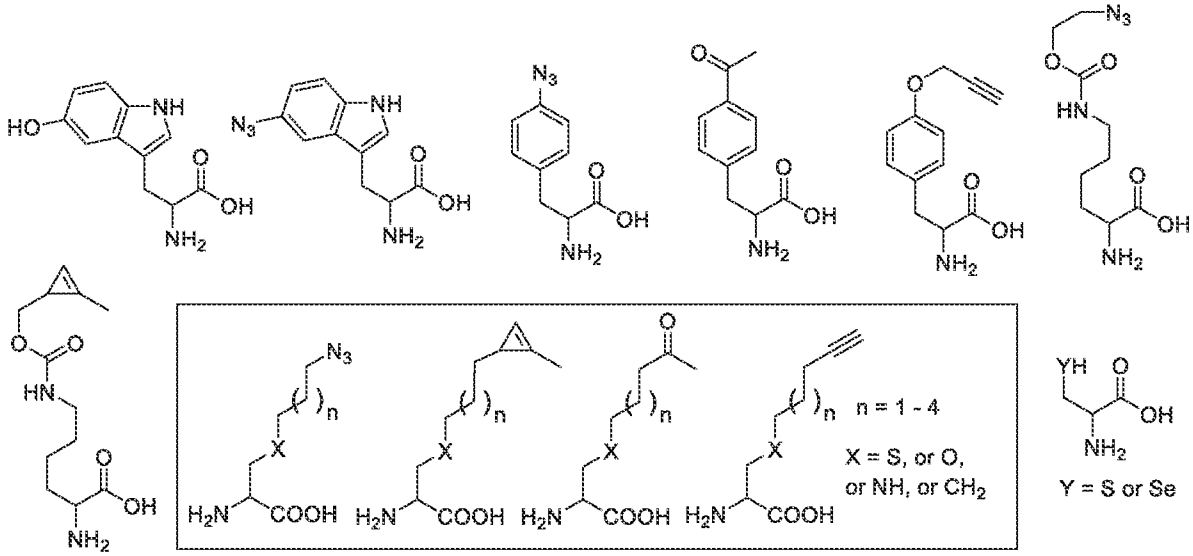
3. The genetically-modified AAV of claim 2, wherein the AAV capsid protein comprises SEQ ID NO:1, or a sequence comprising at least about 80% sequence identity of SEQ ID NO:1.

4. The genetically-modified AAV of claim 3, wherein a stop codon incorporated in the capsid protein(s) and the stop codon is a TAG, TAA or TGA codon.

5. The genetically-modified AAV of claim 1, comprising a variant minor capsid protein, wherein the natural amino acid residue is either a cysteine or a selenocysteine.

6. The genetically-modified AAV of claim 1, comprising a variant minor capsid protein, wherein the unnaturally-occurring amino acid is selected from the group consisting of: phenylalanine analogs; tyrosyl analogs; tryptophanyl analogs; or lysyl analogs.

7. The genetically-modified AAV of claim 6, wherein the analog comprises a formula selected from the group:



8. The genetically-modified AAV of claim 7, comprising a variant minor capsid protein, wherein the analogs are selected from the group consisting of: p-benzoylphenylalanine (pBpA); O-methyltyrosine (OMeY); 5-azidotryptophan; 5-propargyloxytryptopha; 5-aminotryptophan; 5-methoxytryptophan; 5-O-allyltryptophan; 5-bromotryptophan; azido-lysine (AzK); C5Az; LCA; Nε-acetyllysine (AcK); cyclopropene amino acid, *N*$^{\varepsilon}$-(1-methylcycloprop-2-enecarboxamido)-lysine (CpK); 5-hydroxy-tryptophan (5-HTP);LCAlk; DiaazK and LCKet.

9. The genetically-modified AAV of claim 1, wherein the natural or unnatural amino acid residue of the variant minor capsid protein incorporates a bioconjugation handle.

10. The genetically-modified AAV of claim 9, wherein the AAV capsid comprises 5 to 10 bioconjugation handles per capsid.

11. The genetically-modified AAV of any of claims 1 through 10, comprising at least one variant minor capsid protein VP1, VP2 or both VP1 and VP2, wherein the AAV is characterized by:

a) infectivity of target cells comparable to wild-type AAV;

b) is packaged with titers comparable to wild-type AAV; or

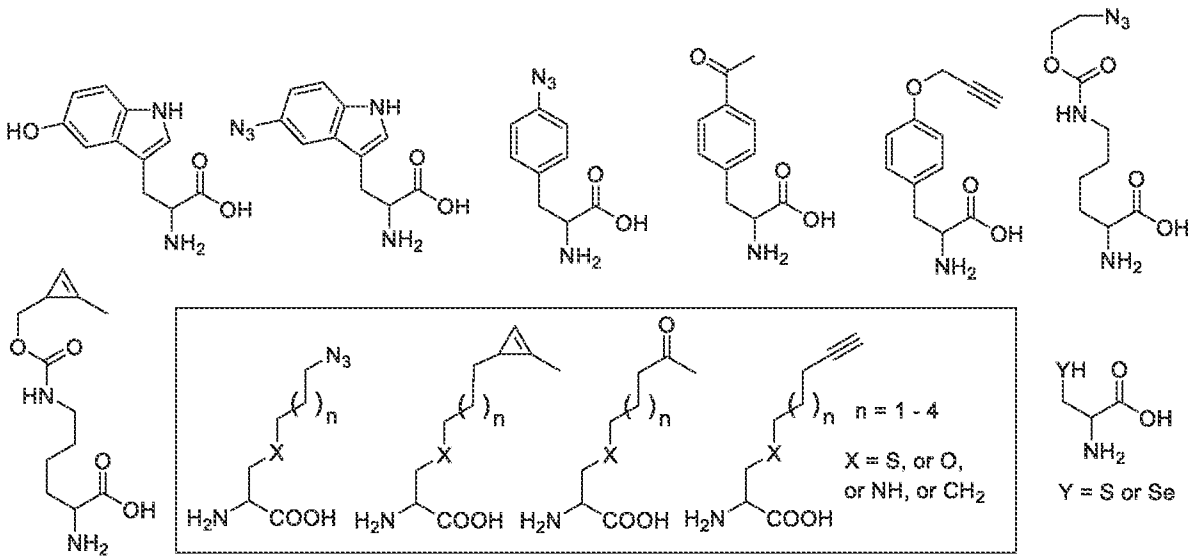
c) both characteristics a) and b).

12. A genetically-modified infectious adeno-associated virus (AAV) wherein the genetically-modified AAV comprises a variant minor capsid protein VP1, VP2 or both VP1 and VP2 wherein VP1, VP2 or both VP1 and VP2 comprise one, or more mutated amino acid residues.

13. The genetically-modified infectious AAV of claim 12, wherein the variant capsid protein comprises one, or more mutated amino acid residues of VP1, VP2 or both VP1 and VP2 and the mutated amino acid residue site incorporates an unnatural amino acid (UAA).

14. The genetically-modified infectious AAV of claim 13, comprising a variant capsid protein, wherein the unnaturally-occurring amino acid is selected from the group consisting of: phenylalanine analogs; tyrosyl analogs; tryptophanyl analogs; or lysyl analogs.

15. The genetically-modified infectious AAV of claim 14, comprising a variant capsid protein, wherein the analogs are selected from the group consisting of:



16. The genetically-modified infectious AAV of any of claims 12 through 15, wherein the variant capsid protein is VP1 comprising SEQ ID NO:1, or a sequence comprising at least about 80% sequence identity of SEQ ID NO:1.

17. The genetically-modified infectious AAV of claim 16, wherein the variant VP1 capsid protein is mutated at one, or more locations of the protein at position(s) 263, 454, 456, 587 and 588.

18. The genetically-modified infectious AAV of any of claims 12 through 17, wherein the variant capsid protein comprises one, or more mutated amino acid residues of VP1,

VP2 or both VP1 and VP2 and the mutated amino acid residue site incorporates a naturally-occurring amino acid.

19. The variant capsid protein of claim 18, wherein the naturally-occurring amino acid is cysteine or selenocysteine.

20. The genetically-modified AAV of claims 18 or 19, wherein the mutated VP1 amino acid sequence comprises SEQ ID NO:1, or a sequence with at least about 80% sequence identity with Seq ID NO:1, wherein the variant VP1 capsid protein is mutated at one, or more locations at position(s) 263, 454, 456, 587 or 588.

21. The genetically-modified infectious AAV of any of claims 12 through 20, comprising the variant capsid protein comprising a mutated amino acid residue, wherein the mutated amino acid is conjugated with a chemical or biological (protein, nucleic acid, lipid, or carbohydrate) entity.

22. The genetically modified infectious AAV comprising a chemical or protein entity of claim 21, wherein the entity is selected from the group consisting of probes, small molecule ligands, peptides, cyclic peptides, nucleotides, polymers, or protein conjugates.

23. The genetically-modified infectious AAV comprising a chemical or protein entity of claim 22, wherein the entity is a cyclic peptide cRGD or polyethylene glycol.

24. A method of producing an infectious genetically-modified AAV, wherein the AAV comprises a variant AAV capsid protein, wherein VP1, VP2 or both VP1 and VP2 comprise one, or more mutated amino acid residue sites, relative to wild-type VP1, VP2 or both VP1 and VP2, the method comprising:

a) providing competent host cells in culture;

b) transfecting the cultured cells with one, or more plasmids comprising:

1) AAV variant VP3 that does not express VP1, or VP2, or both VP1 and VP2;

2) variant VP1,VP2 or both VP1 and VP2 with a suitable promoter that do not express VP3;

3) additional factors required for AAV expression;

c) providing the required unnatural amino acid; and plasmid encoding an engineered aminoacyl-tRNA synthetase/tRNA pair that selectively charge the unnatural amino acid in response to a stop codon;

d) culturing the cells under conditions sufficient for expression of the plasmid genes and assembly of the AAV;

thereby producing an infectious genetically-modified AAV comprising a variant AAV capsid protein, wherein VP1, VP2, VP3 or both VP1 and VP2 are mutated at one, or more amino acid residue sites relative to the wild-type AVV.

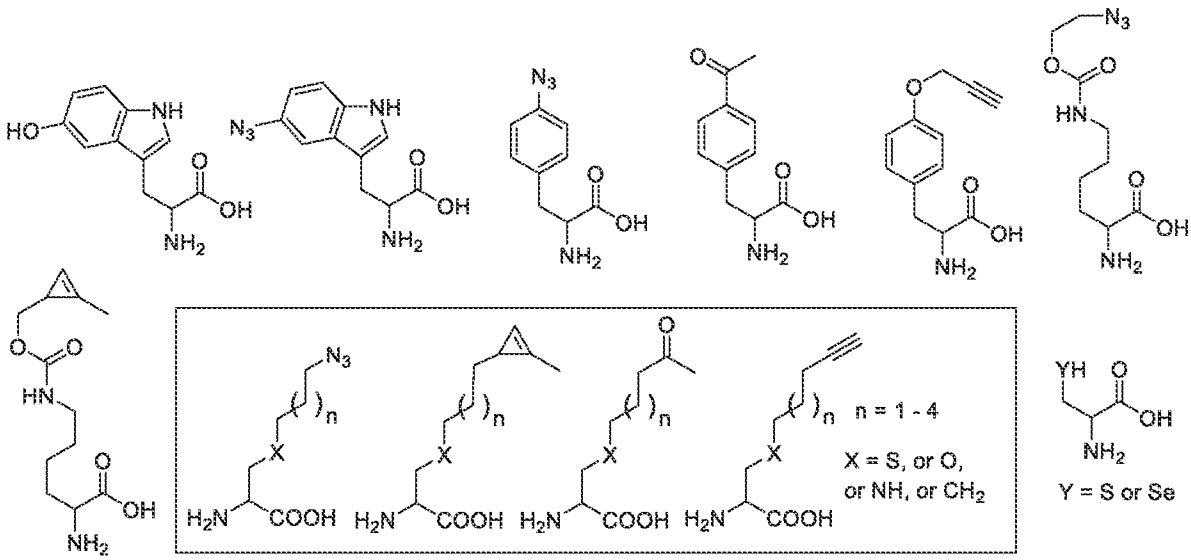
25. The method of claim 24, wherein the suitable promoter is a cytomegalovirus (CMV) promoter.

26. The method of claim 24 or 25, wherein the variant capsid protein is mutated with a naturally-occurring or an unnatural amino acid at the amino acid residue site.

27. The method of claim 25 or 26, wherein the mutated amino acid residue site incorporates an unnatural amino acid (UAA).

28. The method of claim 27, wherein the unnaturally-occurring amino acid is selected from the group consisting of: phenylalanine analogs; tyrosyl analogs; tryptophanyl analogs; or lysyl analogs.

29. The method of claim 28, wherein the analog comprises a formula selected from the group consisting of:



30. The method of claim 24 or 25, wherein the mutated amino acid residue site incorporates a naturally-occurring amino acid.

31. The method of claim 30, wherein the naturally-occurring amino acid is cysteine selenocysteine.

32. The method of any of claims 24 through 31, wherein the mutated amino acid is conjugated with a chemical or biological entity.

33. The method of claim 32, wherein the entity is selected from the group consisting of probes, small molecule ligands, peptides, cyclic peptides, nucleotides, polymers, proteins, or a virus conjugate.

34. The method of claim 33, wherein the entity is a cyclic peptide cRGD or polyethylene glycol.

35. A therapeutic or antigenic composition comprising a genetically-modified AAV of any one of claims 1 through 23 and further comprising one, or more therapeutic or antigenic gene constructs.

36. The composition of claim 35, wherein the composition is an antigenic composition. and the composition further comprises an adjuvant.

37. A method of treating a disease or condition in a subject, the method comprising administering the therapeutic composition of claim 35 to the subject, wherein the composition comprises a gene construct encoding a protein or peptide in a therapeutic amount capable of decreasing or alleviating the disease or condition in the subject.

38. A method of eliciting an immune response in a subject, the method comprising administering the antigenic composition of claim 35 or 36 to the subject, wherein the antigenic composition comprises a gene construct encoding an antigen capable of eliciting an immune response in the subject.

39. A kit comprising the genetically-modified AAV of any of claims 1 through 23.

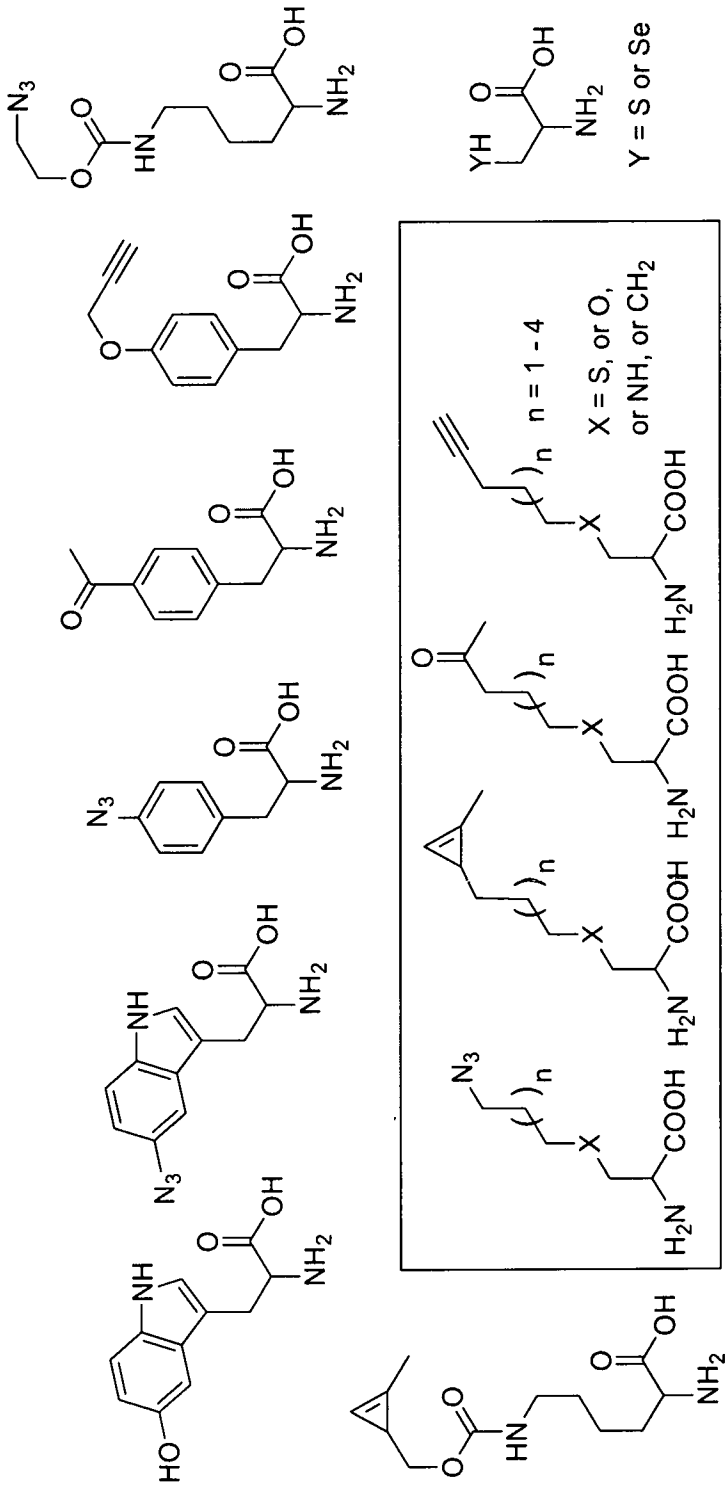


Figure 1

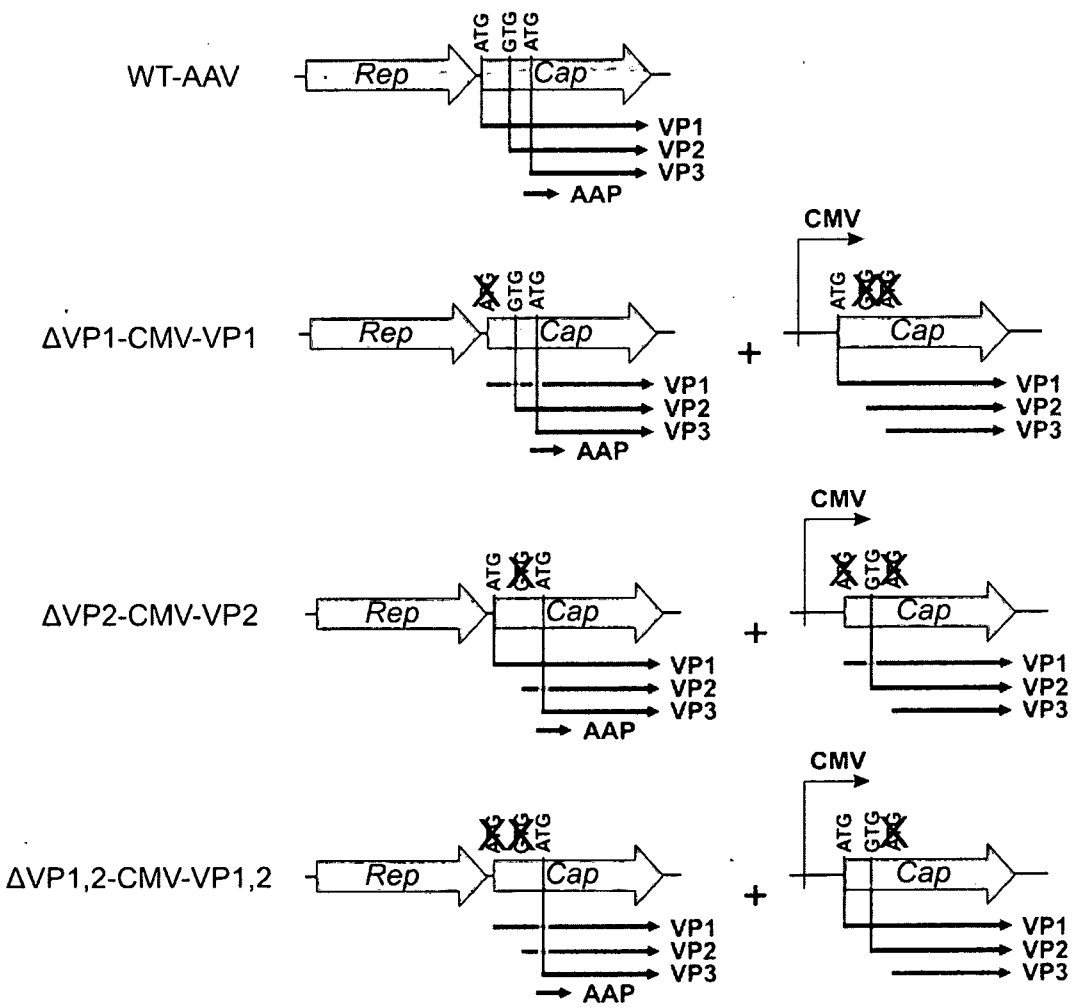

WT-AAV
Rep
Cap
ATG
GTG
ATG
VP1
VP2
VP3
AAP
ΔVP1-CMV-VP1
CMV
ΔVP2-CMV-VP2
ΔVP1,2-CMV-VP1,2
+

Figure 2

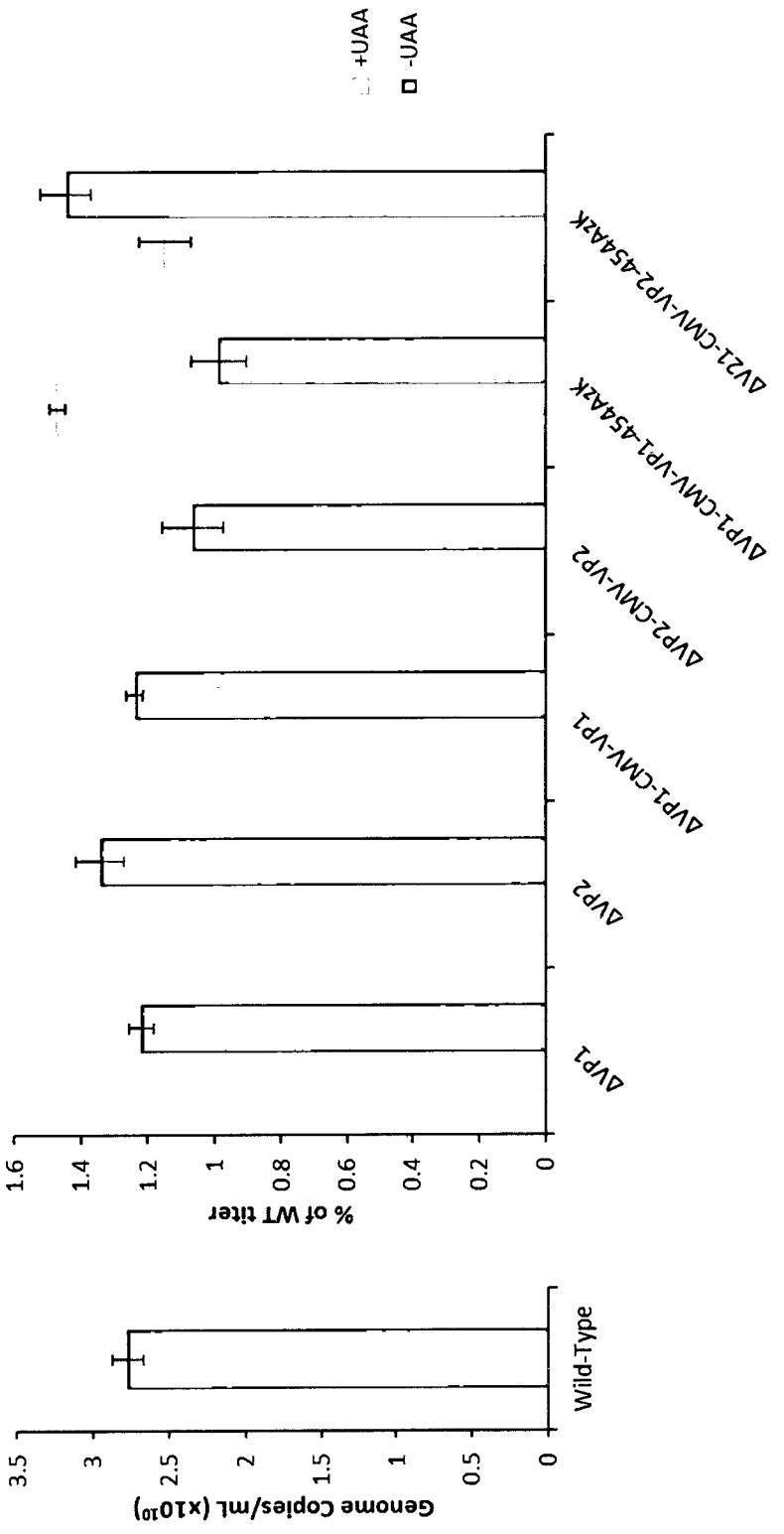

+UAA
-UAA
Genome Copies/mL (x10^10)
3.5
3
2.5
2
1.5
1
0.5
0
Wild-Type
% of WT titer
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
ΔVP1
ΔVP2
ΔVP1-CMV-VP1
ΔVP2-CMV-VP2
ΔVP1-CMV-VP1-454A2K
ΔV21-CMV-VP2-454A2K

Figure 3

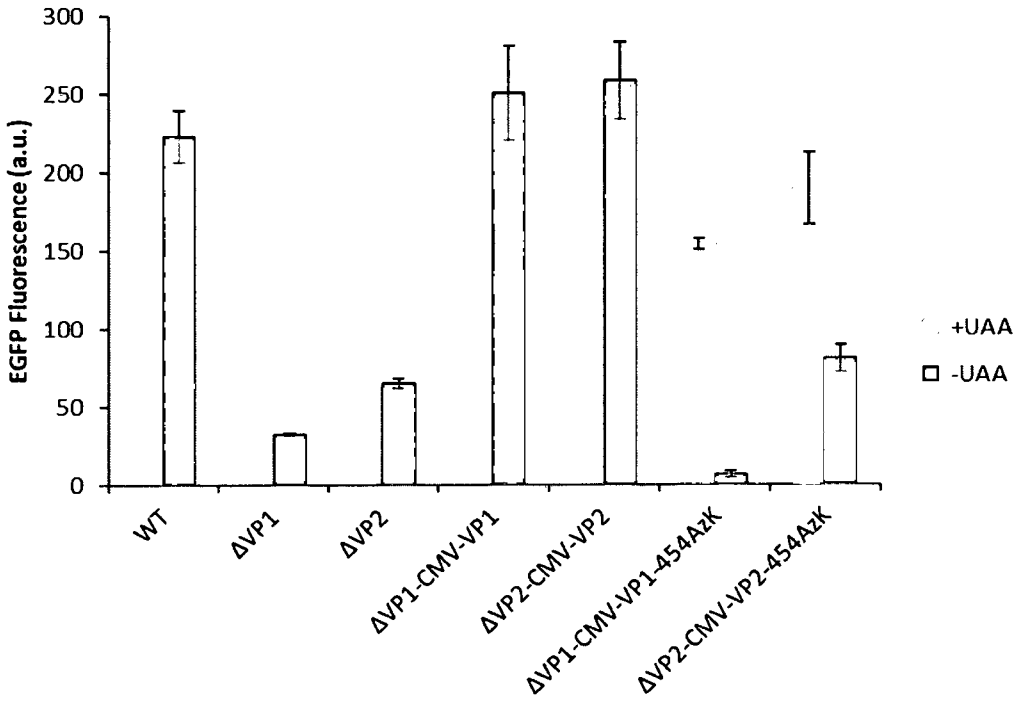

300
250
200
150
100
50
0
EGFP Fluorescence (a.u.)
WT
ΔVP1
ΔVP2
ΔVP1-CMV-VP1
ΔVP2-CMV-VP2
ΔVP1-CMV-VP1-454AzK
ΔVP2-CMV-VP2-454AzK
+UAA
-UAA

Figure 4

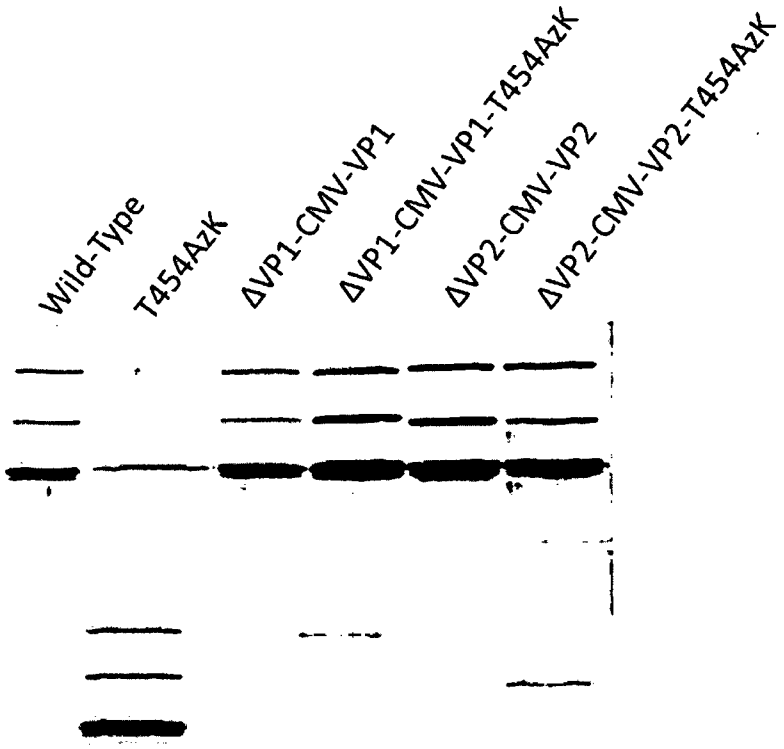

Wild-Type
T454AzK
ΔVP1-CMV-VP1
ΔVP1-CMV-VP1-T454AzK
ΔVP2-CMV-VP2
ΔVP2-CMV-VP2-T454AzK

Figure 5

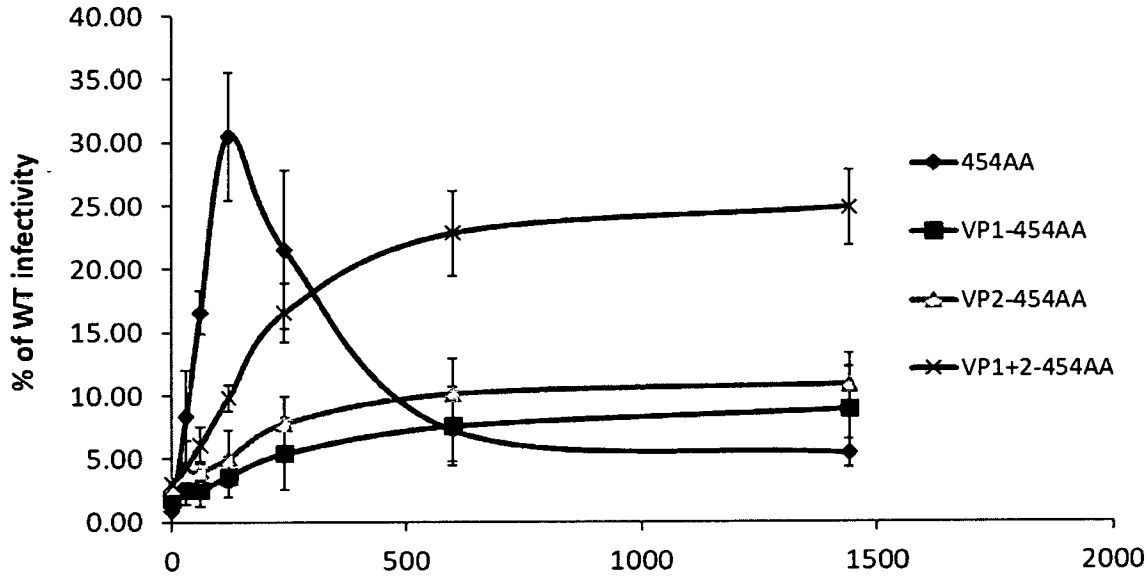

40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
% of WT infectivity
0
500
1000
1500
2000
454AA
VP1-454AA
VP2-454AA
VP1+2-454AA

Figure 6

Figure 7A

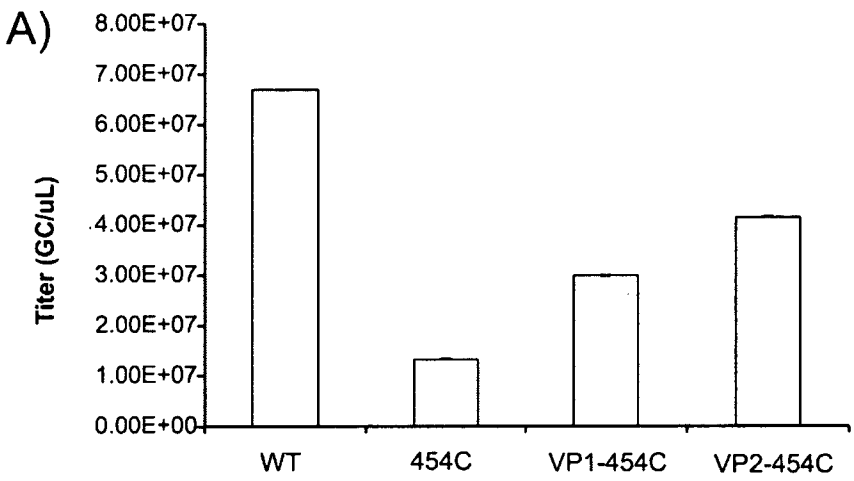


Figure 7B

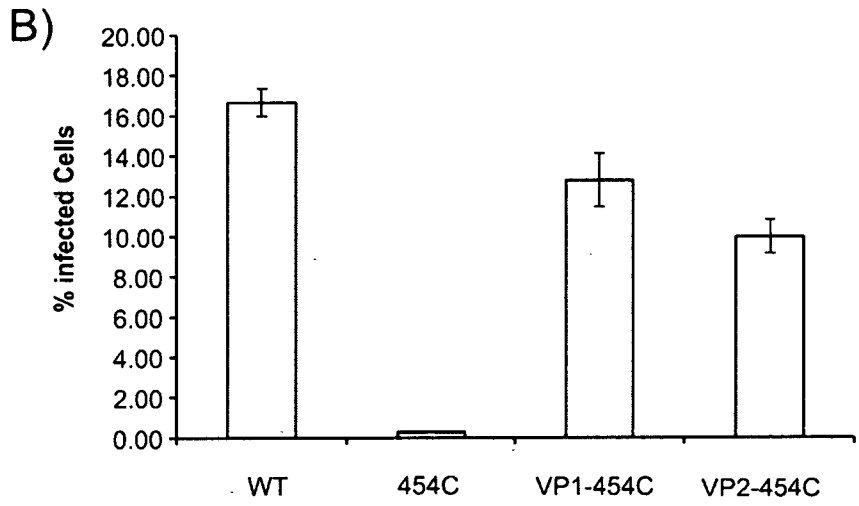
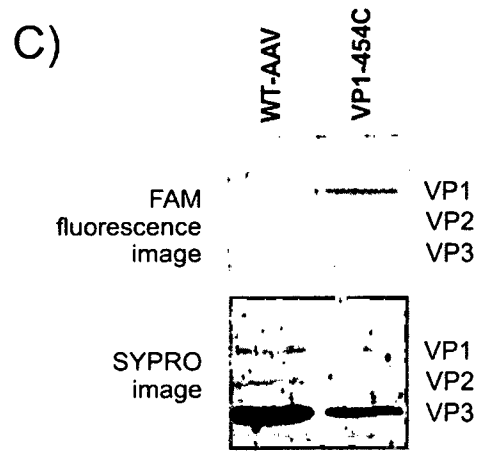


Figure 7C



AAV VP1 amino acid sequence UniProtKB - P03135 (CAPSD_AAV2S)

```
MAADGYLPDW LEDTLSEGIR QWWKLKPGPP PPKPAERHKD DSRGLVLPGY
        60          70          80          90         100
KYLGPFNGLD KGEPVNEADA AALEHDKAYD RQLDSGDNPY LKYNHADAEF
       110         120         130         140         150
QERLKEDTSF GGNLGRAVFQ AKKRVLEPLG LVEEPVKTAP GKKRPVEHSP
       160         170         180         190         200
VEPDSSSGTG KAGQQPARKR LNFGQTGDAD SVPDPQPLGQ PPAAPSGLGT
       210         220         230         240         250
NTMATGSGAP MADNNEGADG VGNSSGNWHC DSTWMGDRVI TTSTRTWALP
       260         270         280         290         300
TYNNHLYKQI SSQSGASNDN HYFGYSTPWG YFDFNRFHCH FSPRDWQRLI
       310         320         330         340         350
NNNWGFRPKR LNFKLFNIQV KEVTQNDGTT TIANNLTSTV QVFTDSEYQL
       360         370         380         390         400
PYVLGSAHQG CLPPFPADVF MVPQYGYLTL NNGSQAVGRS SFYCLEYFPS
       410         420         430         440         450
QMLRTGNNFT FSYTFEDVPF HSSYAHSQSL DRLMNPLIDQ YLYYLSRTNT
       460         470         480         490         500
PSGTTTQSRL QFSQAGASDI RDQSRNWLPG PCYRQQRVSK TSADNNNSEY
       510         520         530         540         550
SWTGATKYHL NGRDSLVNPG PAMASHKDDE EKFFPQSGVL IFGKQGSEKT
       560         570         580         590         600
NVDIEKVMIT DEEEIRTTNP VATEQYGSVS TNLQRGNRQA ATADVNTQGV
       610         620         630         640         650
LPGMVWQDRD VYLQGPIWAK IPHTDGHFHP SPLMGGFGLK HPPPQILIKN
       660         670         680         690         700
TPVPANPSTT FSAAKFASFI TQYSTGQVSV EIEWELQKEN SKRWNPEIQY
       710         720         730
TSNYNKSVNV DFTVDTNGVY SEPRPIGTRY LTRNL
```

Figure 8A

AAV VP2 amino acid sequence UniProtKB - P03135 (CAPSD_AAV2S)

```
MAPGKKRPVE HSPVEPDSSS GTGKAGQQPA RKRLNFGQTG DADSVPDPQP
        60          70          80          90         100
LGQPPAAPSG LGTNTMATGS GAPMADNNEG ADGVGNSSGN WHCDSTWMGD
       110         120         130         140         150
RVITTSTRTW ALPTYNNHLY KQISSQSGAS NDNHYFGYST PWGYFDFNRF
       160         170         180         190         200
HCHFSPRDWQ RLINNNWGFR PKRLNFKLFN IQVKEVTQND GTTTIANNLT
       210         220         230         240         250
STVQVFTDSE YQLPYVLGSA HQGCLPPFPA DVFMVPQYGY LTLNNGSQAV
       260         270         280         290         300
GRSSFYCLEY FPSQMLRTGN NFTFSYTFED VPFHSSYAHS QSLDRLMNPL
       310         320         330         340         350
IDQYLYYLSR TNTPSGTTTQ SRLQFSQAGA SDIRDQSRNW LPGPCYRQQR
       360         370         380         390         400
VSKTSADNNN SEYSWTGATK YHLNGRDSLV NPGPAMASHK DDEEKFFPQS
       410         420         430         440         450
GVLIFGKQGS EKTNVDIEKV MITDEEEIRT TNPVATEQYG SVSTNLQRGN
       460         470         480         490         500
RQAATADVNT QGVLPGMVWQ DRDVYLQGPI WAKIPHTDGH FHPSPLMGGF
       510         520         530         540         550
GLKHPPPQIL IKNTPVPANP STTFSAAKFA SFITQYSTGQ VSVEIEWELQ
       560         570         580         590
KENSKRWNPE IQYTSNYNKS VNVDFTVDTN GVYSEPRPIG TRYLTRNL
```

Figure 8B

AAV VP3 amino acid sequence UniProtKB - P03135 (CAPSD_AAV2S)

```
MATGSGAPMA DNNEGADGVG NSSGNWHCDS TWMGDRVITT STRTWALPTY
        60          70          80          90         100
NNHLYKQISS QSGASNDNHY FGYSTPWGYF DFNRFHCHFS PRDWQRLINN
       110         120         130         140         150
NWGFRPKRLN FKLFNIQVKE VTQNDGTTTI ANNLTSTVQV FTDSEYQLPY
       160         170         180         190         200
VLGSAHQGCL PPFPADVFMV PQYGYLTLNN GSQAVGRSSF YCLEYFPSQM
       210         220         230         240         250
LRTGNNFTFS YTFEDVPFHS SYAHSQSLDR LMNPLIDQYL YYLSRTNTPS
       260         270         280         290         300
GTTTQSRLQF SQAGASDIRD QSRNWLPGPC YRQQRVSKTS ADNNNSEYSW
       310         320         330         340         350
TGATKYHLNG RDSLVNPGPA MASHKDDEEK FFPQSGVLIF GKQGSEKTNV
       360         370         380         390         400
DIEKVMITDE EEIRTTNPVA TEQYGSVSTN LQRGNRQAAT ADVNTQGVLP
       410         420         430         440         450
GMVWQDRDVY LQGPIWAKIP HTDGHFHPSP LMGGFGLKHP PPQILIKNTP
       460         470         480         490         500
VPANPSTTFS AAKFASFITQ YSTGQVSVEI EWELQKENSK RWNPEIQYTS
       510         520         530
NYNKSVNVDF TVDTNGVYSE PRPIGTRYLT RNL
```

Figure 8C

AAV VP2 amino acid sequence UniProtKB - P03135 (CAPSD_AAV2S)

```
MAPGKKRPVE HSPVEPDSSS GTGKAGQQPA RKRLNFGQTG DADSVPDPQP
        60          70          80          90         100
LGQPPAAPSG LGTNTMATGS GAPMADNNEG ADGVGNSSGN WHCDSTWMGD
       110         120         130         140         150
RVITTSTRTW ALPTYNNHLY KQISSQSGAS NDNHYFGYST PWGYFDFNRF
       160         170         180         190         200
HCHFSPRDWQ RLINNNWGFR PKRLNFKLFN IQVKEVTQND GTTTIANNLT
       210         220         230         240         250
STVQVFTDSE YQLPYVLGSA HQGCLPPFPA DVFMVPQYGY LTLNNGSQAV
       260         270         280         290         300
GRSSFYCLEY FPSQMLRTGN NFTFSYTFED VPFHSSYAHS QSLDRLMNPL
       310         320         330         340         350
IDQYLYYLSR TNTPSGTTTQ SRLQFSQAGA SDIRDQSRNW LPGPCYRQQR
       360         370         380         390         400
VSKTSADNNN SEYSWTGATK YHLNGRDSLV NPGPAMASHK DDEEKFFPQS
       410         420         430         440         450
GVLIFGKQGS EKTNVDIEKV MITDEEEIRT TNPVATEQYG SVSTNLQRGN
       460         470         480         490         500
RQAATADVNT QGVLPGMVWQ DRDVYLQGPI WAKIPHTDGH FHPSPLMGGF
       510         520         530         540         550
GLKHPPPQIL IKNTPVPANP STTFSAAKFA SFITQYSTGQ VSVEIEWELQ
       560         570         580         590
KENSKRWNPE IQYTSNYNKS VNVDFTVDTN GVYSEPRPIG TRYLTRNL
```

Figure 9

AAV VP3 amino acid sequence UniProtKB - P03135 (CAPSD_AAV2S)

```
MATGSGAPMA DNNEGADGVG NSSGNWHCDS TWMGDRVITT STRTWALPTY
        60          70          80          90         100
NNHLYKQISS QSGASNDNHY FGYSTPWGYF DFNRFHCHFS PRDWQRLINN
       110         120         130         140         150
NWGFRPKRLN FKLFNIQVKE VTQNDGTTTI ANNLTSTVQV FTDSEYQLPY
       160         170         180         190         200
VLGSAHQGCL PPFPADVFMV PQYGYLTLNN GSQAVGRSSF YCLEYFPSQM
       210         220         230         240         250
LRTGNNFTFS YTFEDVPFHS SYAHSQSLDR LMNPLIDQYL YYLSRTNTPS
       260         270         280         290         300
GTTTQSRLQF SQAGASDIRD QSRNWLPGPC YRQQRVSKTS ADNNNSEYSW
       310         320         330         340         350
TGATKYHLNG RDSLVNPGPA MASHKDDEEK FFPQSGVLIF GKQGSEKTNV
       360         370         380         390         400
DIEKVMITDE EEIRTTNPVA TEQYGSVSTN LQRGNRQAAT ADVNTQGVLP
       410         420         430         440         450
GMVWQDRDVY LQGPIWAKIP HTDGHFHPSP LMGGFGLKHP PPQILIKNTP
       460         470         480         490         500
VPANPSTTFS AAKFASFITQ YSTGQVSVEI EWELQKENSK RWNPEIQYTS
       510         520         530
NYNKSVNVDF TVDTNGVYSE PRPIGTRYLT RNL
```

Figure 10

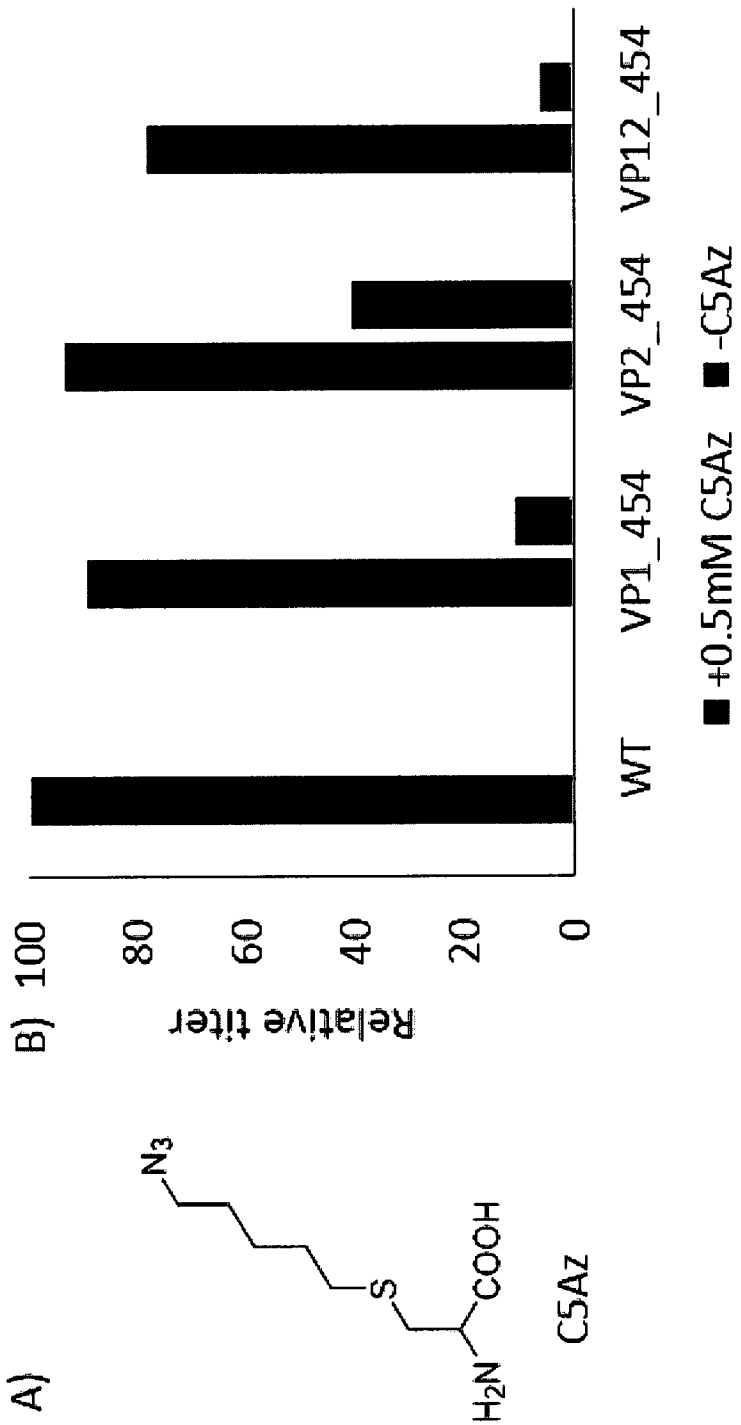


Figure 11A

Figure 11B

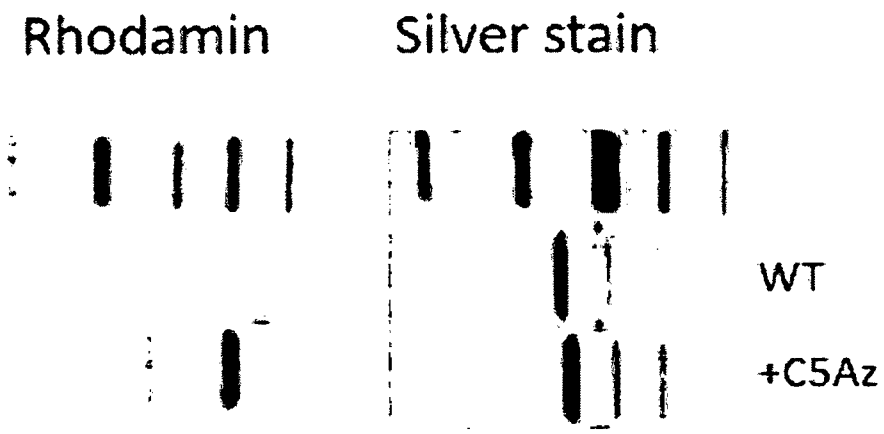


Figure 11C

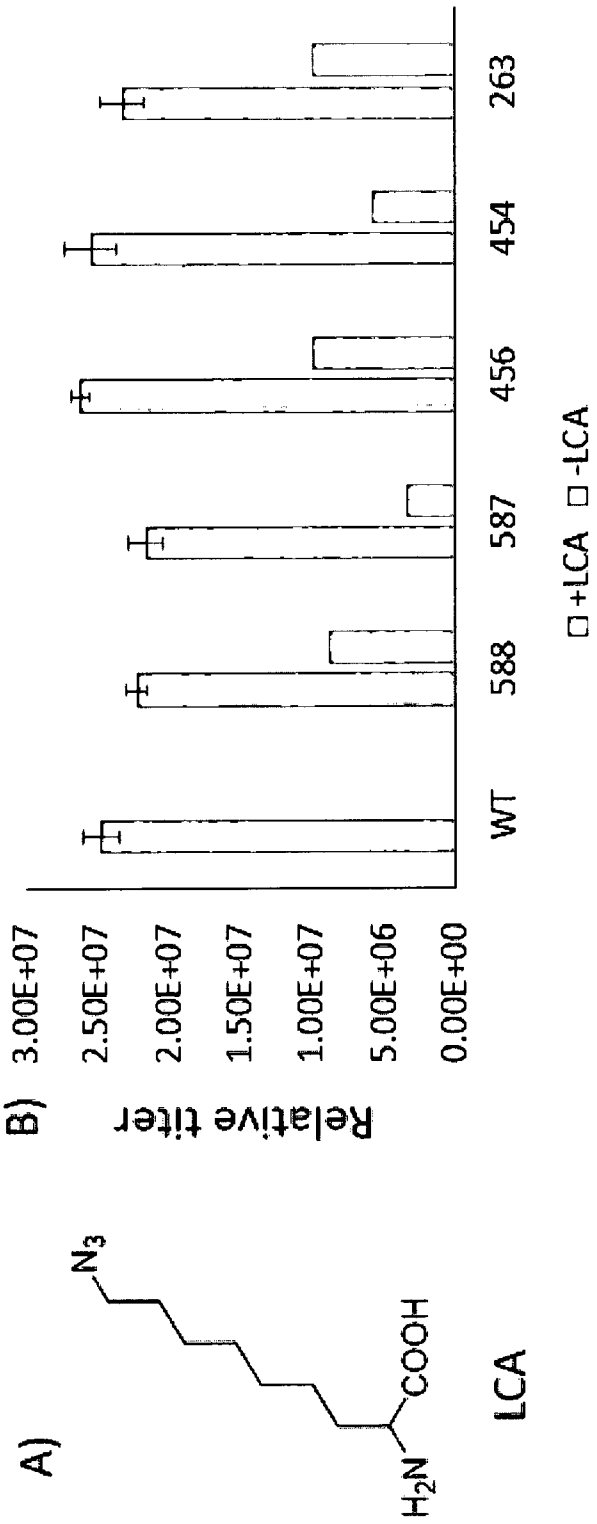


Figure 12A

Figure 12B

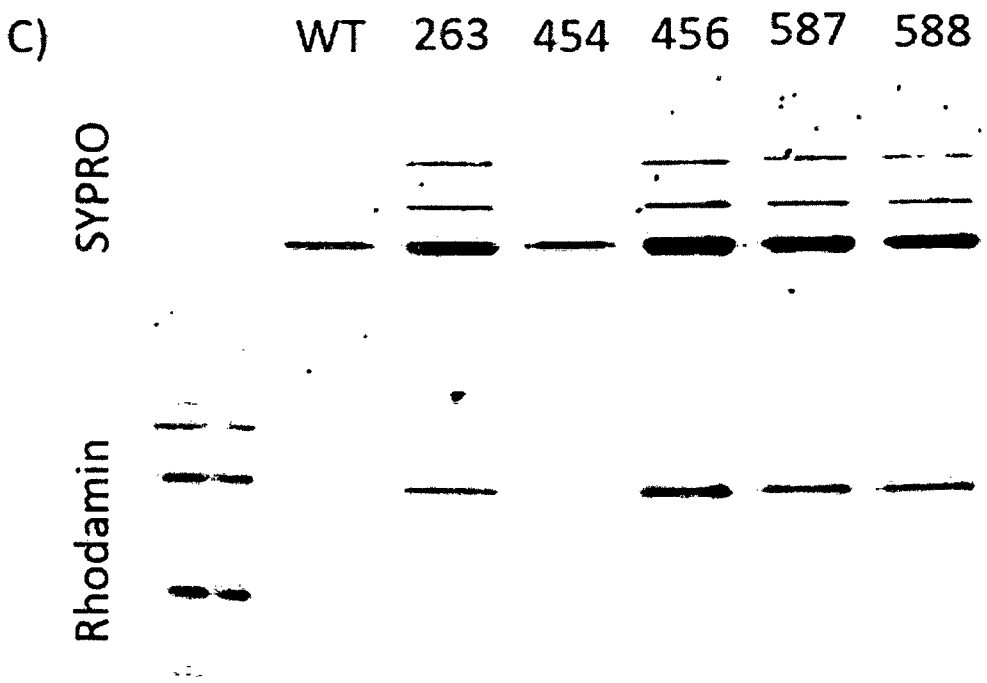


Figure 12C

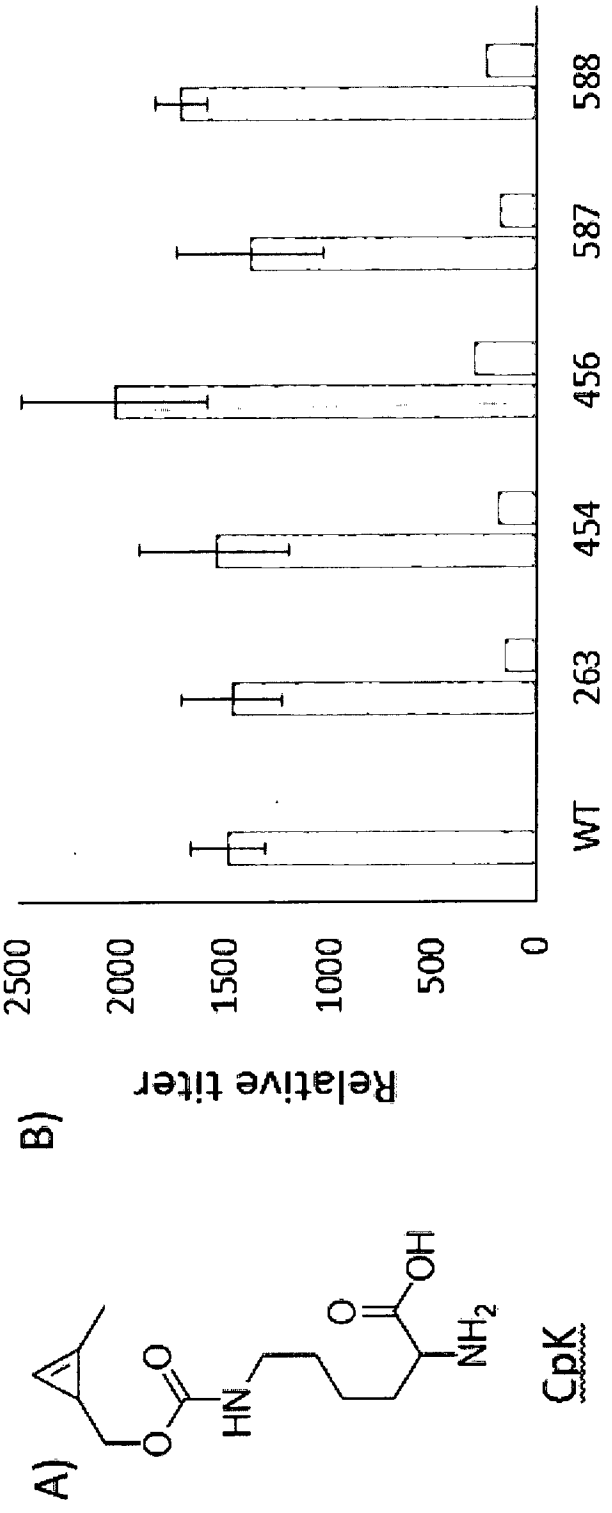


Figure 13A

Figure 13B

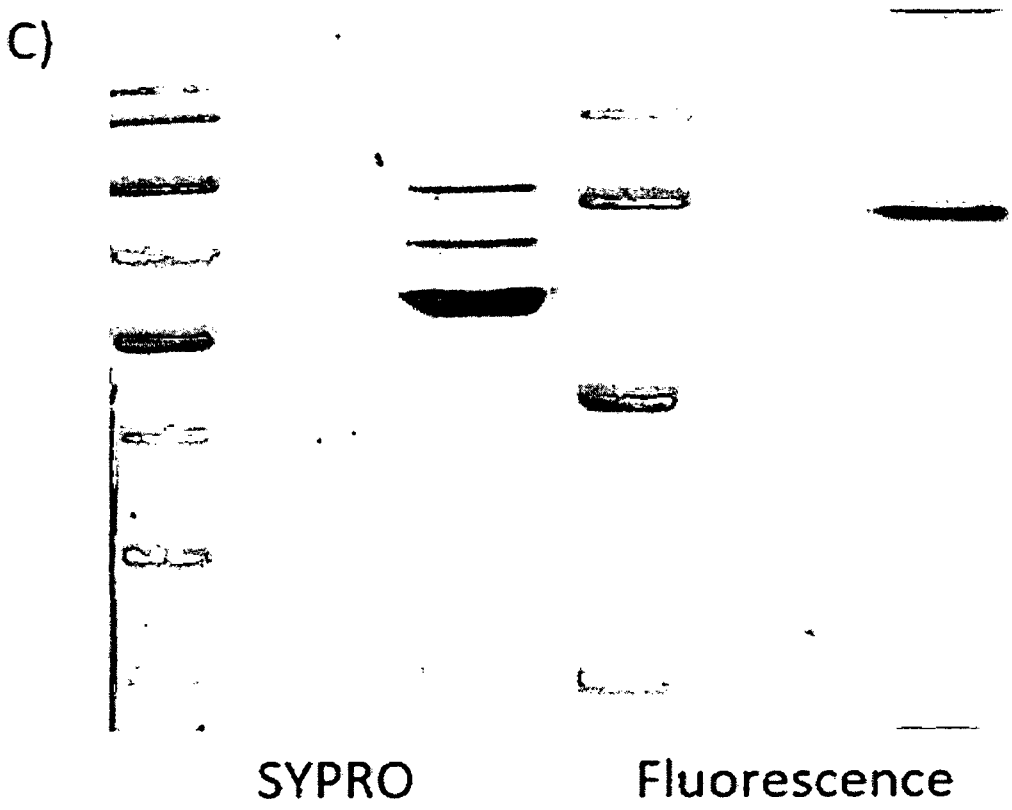


Figure 13C

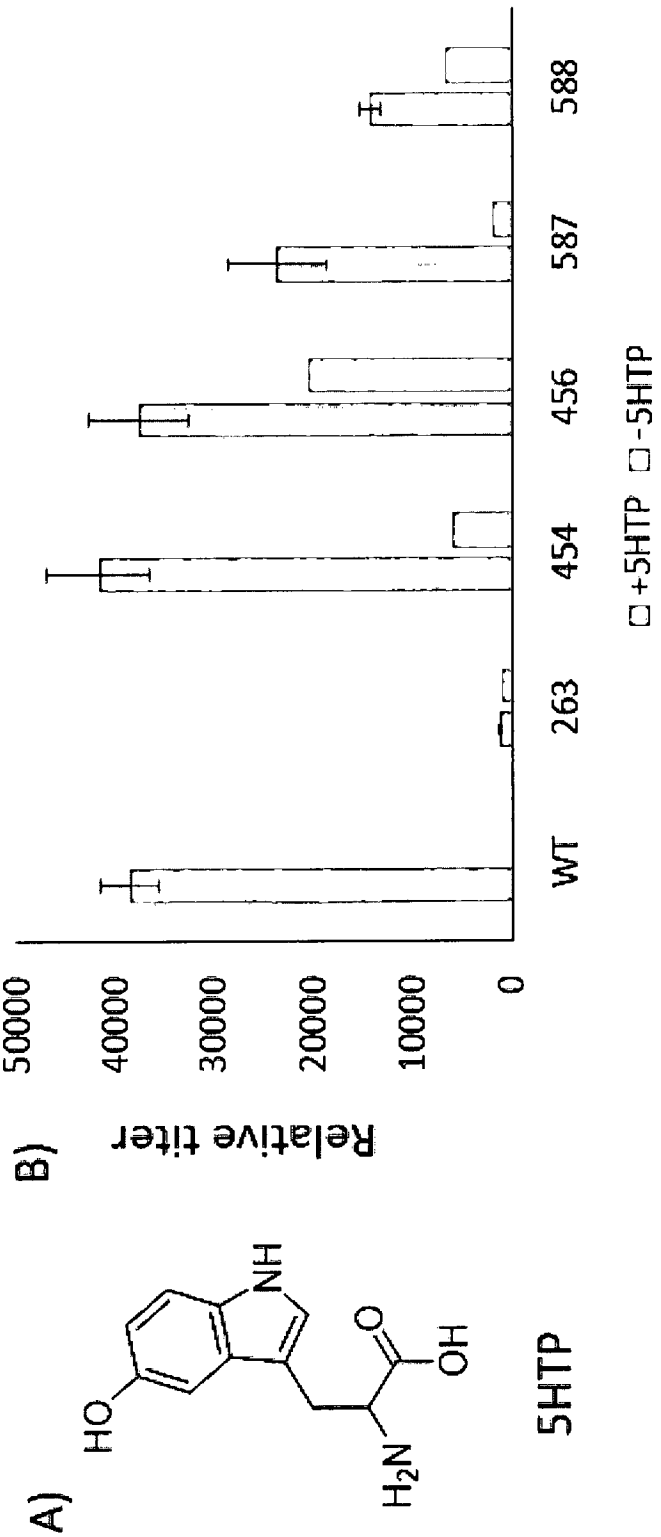


Figure 14A

Figure 14B

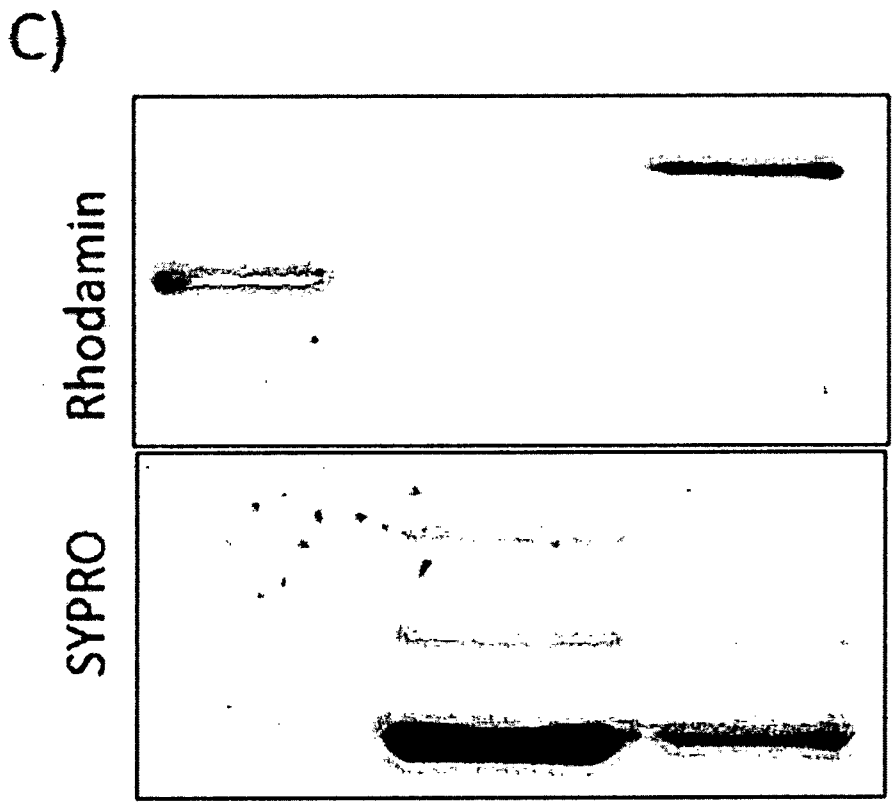


Figure 14C

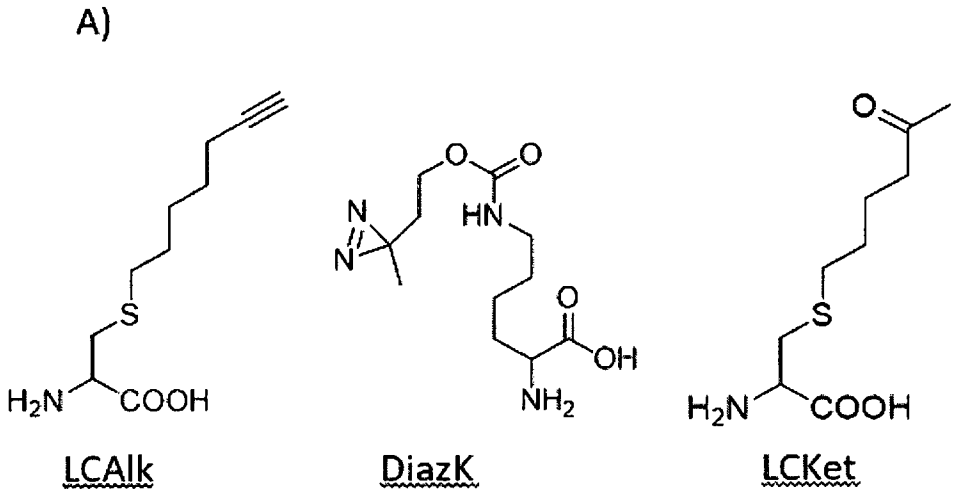


Figure 15A

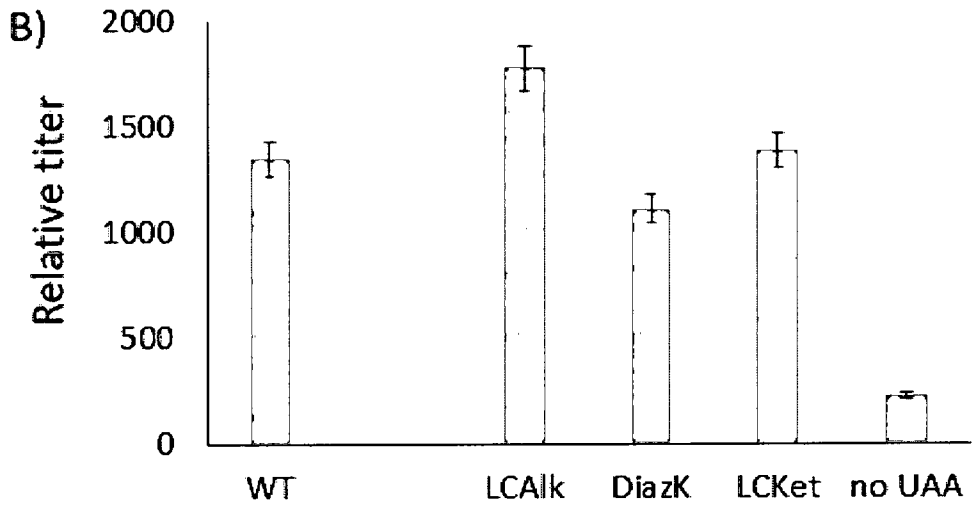


Figure 15B

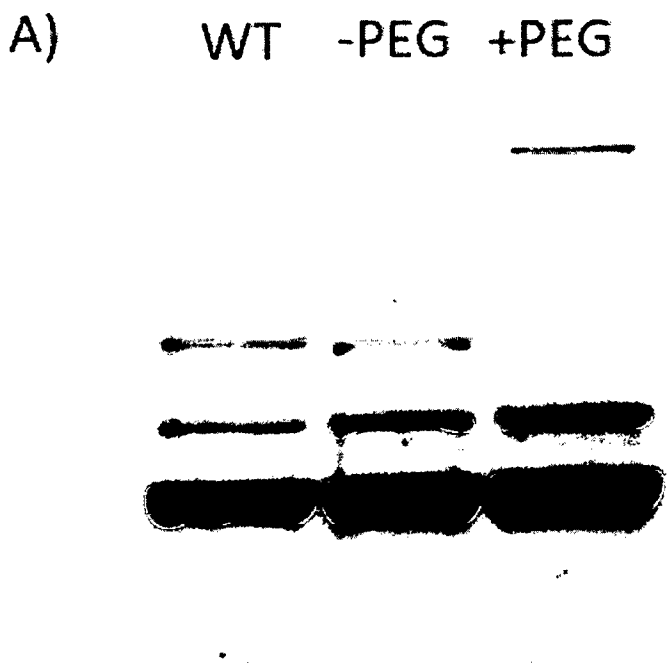


Figure 16A

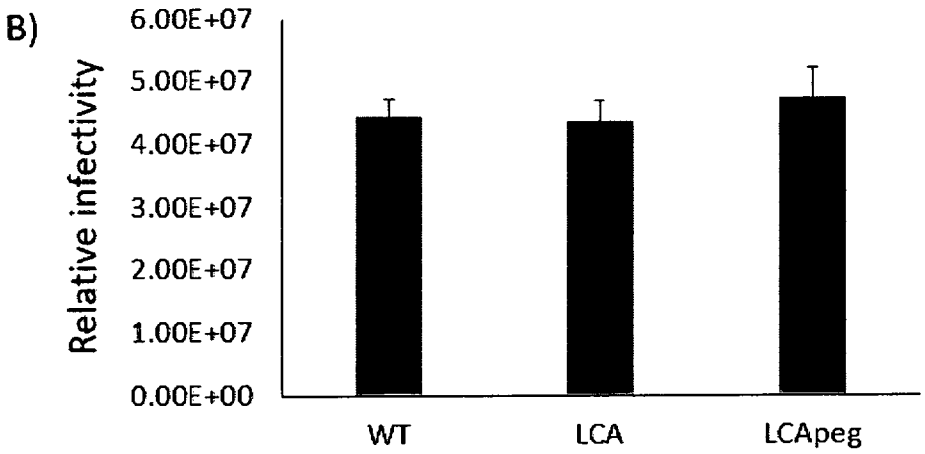


Figure 16B

RC2-VP1-del:

gctctagaactagtggatcccccggaagatcagaagttcctattccgaagttcctattctctagaaagtataggaacttctgatctgc
gcagccgccatgccggggttttacgagattgtgattaaggtccccagcgaccttgacgagcatctgcccggcatttctgacagcttt
gtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatctgattgagcaggcacccctgacc
gtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctttttctttgtgcaatttga
gaagggagagagctacttccacatgcacgtgctcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcag
attcgcgaaaaactgattcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaat
ggcgccggaggcgggaacaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctgagctccagtg
ggcgtggactaatatggaacagtatttaagcgcctgtttgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcac
gtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaattctgatgcgccggtgatcagatcaaaaacttcagcca
ggtacatggagctggtcgggtggctcgtggacaaggggattacctcggagaagcagtggatccaggaggaccaggcctcata
catctccttcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaa
accgcccccgactacctggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaattttggaactaaacgggt
acgatccccaatatgcggcttccgtctttctgggatgggccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcct
gcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgcccttctacgggtgcgtaaactggaccaatgaga
actttcccttcaacgactgtgtcgacaagatggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggcca
aagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgt
cacctccaacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagaccggatgt
tcaaatttgaactcacccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagacttttccggtgggcaaa
ggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtggagccaagaaaagacccgcccccagtgacgcagat
ataagtgagcccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagaca
ggtaccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattca
aatatctgcttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgtttctgtcgtcaaaaaggcgt
atcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatctggtcaatgtggatttggatg
actgcatctttgaacaataaatgatttaaatcaggtCTCgctgccgatggttatcttccagattggctcgaggacactctctctgaa
ggaataagacagtggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggacgacagcaggg
gtcttgtgcttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccggtcaacgaggcagacgccgcg
gccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacctcaagtacaaccacgccgacgc
ggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtcttccaggcgaaaaagagggttctt
gaacctctgggcctggttgaggaacctgttaagacggctccgggaaaaaagaggccggtagagcactctcctgtggagccag
actcctcctcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaattttggtcagactggagacgcagactc
agtacctgacccccagcctctcggacagccaccagcagccccctctggtctgggaactaatacgatggctacaggcagtggcg
caccaatggcagacaataacgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatggatgggc
gacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaaccacctctacaaacaaatttccagccaat
caggagcctcgaacgacaatcactactttggctacagcacccttgggggtattttgacttcaacagattccactgccacttttcacc
acgtgactggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaa
agaggtcacgcagaatgacggtacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtacca
gctcccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgccacagtatggatacc
tcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagtactttccttctcagatgctgcgtaccggaa
acaactttaccttcagctacacttttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtctcatgaatcc
tctcatcgaccagtacctgtattacttgagcagaacaaacactccaagtggaaccaccacgcagtcaaggcttcagttttctcagg
ccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaaagacatctg
cggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggcagagactctctggtgaatccgggc
ccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggggttctcatctttgggaagcaaggctcagag

Figure 17

aaaacaaatgtggacattgaaaaggtcatgattacagacgaagaggaaatcaggacaaccaatcccgtggctacggagcag
tatggttctgtatctaccaacctccagagaggcaacagacaagcagctaccgcagatgtcaacacacaaggcgttcttccaggc
atggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagattccacacacggacggacattttcacccctctc
ccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaacaccccggtacctgcgaatccttcgaccacc
ttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtcagcgtggagatcgagtgggagctgcagaag
gaaaacagcaaacgctggaatcccgaaattcagtacacttccaactacaacaagtctgttaatgtggactttactgtggacacta
atggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaatctgtaattgcttgttaatcaataaaccgtttaattcg
tttcagttgaactttggtctctgcgtatttctttcttatctagtttccatggctacgtagataagtagcatggcgggttaatcattaactacag
cccgggcgtttaaacagcgggcggaggggtggagtcgtgacgtgaattacgtcatagggttagggaggtcctgtattagaggtc
acgtgagtgttttgcgacattttgcgacaccatgtggtctcgctgggggggggggcccgagtgagcacgcagggtctccattttga
agcgggaggtttgaacgagcgctggcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaa
cttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttccc

Figure 17 continued

RC2-VP2-del:

gctctagaactagtggatcccccggaagatcagaagttcctattccgaagttcctattctctagaaagtataggaacttctgatctgc
gcagccgccatgccggggttttacgagattgtgattaaggtccccagcgaccttgacgagcatctgcccggcatttctgacagcttt
gtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatctgattgagcaggcacccctgacc
gtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctttttctttgtgcaatttga
gaagggagagagctacttccacatgcacgtgctcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcag
attcgcgaaaaactgattcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaat
ggcgccggaggcgggaacaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctgagctccagtg
ggcgtggactaatatggaacagtatttaagcgcctgtttgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcac
gtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaattctgatgcgccggtgatcagatcaaaaacttcagcca
ggtacatggagctggtcgggtggctcgtggacaaggggattacctcggagaagcagtggatccaggaggaccaggcctcata
catctccttcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaa
accgcccccgactacctggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaatttggaactaaacgggt
acgatccccaatatgcggcttccgtctttctgggatgggccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcct
gcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgcccttctacgggtgcgtaaactggaccaatgaga
actttcccttcaacgactgtgtcgacaagatggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggcca
aagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgt
cacctccaacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagaccggatgt
tcaaatttgaactcacccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagacttttccggtgggcaaa
ggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtggagccaagaaaagacccgcccccagtgacgcagat
ataagtgagcccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagaca
ggtaccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattca
aatatctgcttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgtttctgtcgtcaaaaaggcgt
atcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatctggtcaatgtggatttggatg
actgcatctttgaacaataaatgatttaaatcaggtatggctgccgatggttatcttccagattggctcgaggacactctctctgaagg
aataagacagtggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggacgacagcaggggt
cttgtgcttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccggtcaacgaggcagacgccgcgg
ccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacctcaagtacaaccacgccgacgcg
gagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtcttccaggcgaaaaagagggttcttg
aacctctgggcctggttgaggaacctgttaagACCgctccgggaaaaaagaggccggtagagcactctcctgtggagccag
actcctcctcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaattttggtcagactggagacgcagactc
agtacctgacccccagcctctcggacagccaccagcagccccctctggtctgggaactaatacgatggctacaggcagtggcg
caccaatggcagacaataacgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatggatgggc
gacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaaccacctctacaaacaaatttccagccaat
caggagcctcgaacgacaatcactactttggctacagcacccttggggtattttgacttcaacagattccactgccacttttcacc
acgtgactggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaa
agaggtcacgcagaatgacggtacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtacca
gctcccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgccacagtatggatacc
tcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagtactttccttctcagatgctgcgtaccggaa
acaactttaccttcagctacacttttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtctcatgaatcc
tctcatcgaccagtacctgtattacttgagcagaacaaacactccaagtggaaccaccacgcagtcaaggcttcagttttctcagg
ccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaaagacatctg
cggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggcagagactctctggtgaatccgggc
ccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggggttctcatctttgggaagcaaggctcagag

Figure 18

aaaacaaatgtggacattgaaaaggtcatgattacagacgaagaggaaatcaggacaaccaatcccgtggctacggagcag
tatggttctgtatctaccaacctccagagaggcaacagacaagcagctaccgcagatgtcaacacacaaggcgttcttccaggc
atggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagattccacacacggacggacattttcacccctctc
ccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaacaccccggtacctgcgaatccttcgaccacc
ttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtcagcgtggagatcgagtgggagctgcagaag
gaaaacagcaaacgctggaatcccgaaattcagtacacttccaactacaacaagtctgttaatgtggactttactgtggacacta
atggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaatctgtaattgcttgttaatcaataaaccgtttaattcg
tttcagttgaactttggtctctgcgtatttctttcttatctagtttccatggctacgtagataagtagcatggcgggttaatcattaactacag
cccgggcgtttaaacagcgggcggaggggtggagtcgtgacgtgaattacgtcatagggttagggaggtcctgtattagaggtc
acgtgagtgttttgcgacattttgcgacaccatgtggtctcgctgggggggggggcccgagtgagcacgcagggtctccattttga
agcgggaggtttgaacgagcgctggcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaa
cttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttccc

Figure 18 continued

RC2-VP12-del:

gctctagaactagtggatcccccggaagatcagaagttcctattccgaagttcctattctctagaaagtataggaacttctgatctgc
gcagccgccatgccgggggttttacgagattgtgattaaggtccccagcgaccttgacgagcatctgcccggcatttctgacagcttt
gtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatctgattgagcaggcacccctgacc
gtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctttttctttgtgcaatttga
gaagggagagagctacttccacatgcacgtgctcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcag
attcgcgaaaaactgattcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaat
ggcgccggaggcgggaacaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctgagctccagtg
ggcgtggactaatatggaacagtatttaagcgcctgtttgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcac
gtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaattctgatgcgccggtgatcagatcaaaaacttcagcca
ggtacatggagctggtcgggtggctcgtggacaaggggattacctcggagaagcagtggatccaggaggaccaggcctcata
catctccttcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaa
accgcccccgactacctggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaattttggaactaaacgggt
acgatccccaatatgcggcttccgtctttctgggatgggccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcct
gcaactaccgggaagaccaacatcgcggaggccatagcccacactgtgcccttctacgggtgcgtaaactggaccaatgaga
actttcccttcaacgactgtgtcgacaagatggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtcggcca
aagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgt
cacctccaacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagaccggatgt
tcaaatttgaactcacccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagacttttccggtgggcaaa
ggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtggagccaagaaaagacccgcccccagtgacgcagat
ataagtgagcccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcggaagcttcgatcaactacgcagaca
ggtaccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttccctgcagacaatgcgagagaatgaatcagaattca
aatatctgcttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgtttctgtcgtcaaaaaggcgt
atcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatctggtcaatgtggatttggatg
actgcatctttgaacaataaatgatttaaatcaggtCTCgctgccgatggttatcttccagattggctcgaggacactctctctgaa
ggaataagacagtggtggaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggacgacagcaggg
gtcttgtgcttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccggtcaacgaggcagacgccgcg
gccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacctcaagtacaaccacgccgacgc
ggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtcttccaggcgaaaaagagggttctt
gaacctctgggcctggttgaggaacctgttaagACCgctccgggaaaaaagaggccggtagagcactctcctgtggagcca
gactcctcctcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaattttggtcagactggagacgcagact
cagtacctgacccccagcctctcggacagccaccagcagccccctctggtctgggaactaatacgatggctacaggcagtggc
gcaccaatggcagacaataacgagggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatggatggg
cgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaaccacctctacaaacaaatttccagccaa
tcaggagcctcgaacgacaatcactactttggctacagcacccttgggggtattttgacttcaacagattccactgccacttttcac
cacgtgactggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctctttaacattcaagtca
aagaggtcacgcagaatgacggtacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtacc
agctcccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgccacagtatggata
cctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagtactttccttctcagatgctgcgtaccgg
aaacaactttaccttcagctacacttttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtctcatgaat
cctctcatcgaccagtacctgtattacttgagcagaacaaacactccaagtggaaccaccacgcagtcaaggcttcagttttctca
ggccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaaagacat
ctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggcagagactctctggtgaatccgg
gcccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggggttctcatctttgggaagcaaggctcag

Figure 19

agaaaacaaatgtggacattgaaaaggtcatgattacagacgaagaggaaatcaggacaaccaatcccgtggctacggagc
agtatggttctgtatctaccaacctccagagaggcaacagacaagcagctaccgcagatgtcaacacacaaggcgttcttccag
gcatggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagattccacacacggacggacattttcacccct
ctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaacaccccggtacctgcgaatccttcgacc
accttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtcagcgtggagatcgagtgggagctgcaga
aggaaaacagcaaacgctggaatcccgaaattcagtacacttccaactacaacaagtctgttaatgtggactttactgtggacac
taatggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaatctgtaattgcttgttaatcaataaaccgtttaatt
cgtttcagttgaactttggtctctgcgtatttctttcttatctagtttccatggctacgtagataagtagcatggcgggttaatcattaactac
agcccgggcgtttaaacagcgggcggaggggtggagtcgtgacgtgaattacgtcatagggttagggaggtcctgtattagagg
tcacgtgagtgttttgcgacattttgcgacaccatgtggtctcgctgggggggggggcccgagtgagcacgcagggtctccattttg
aagcgggaggtttgaacgagcgctggcgcgctcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttaccca
acttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttccc

Figure 19 continued

CMV-VP1-delVP23:

gttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactt
acggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgcca
atagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagta
cgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagt
acatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgggg
atttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactcc
gccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccact
gcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcATGgctgccgatggttatcttccagatt
ggctcgaggacactctctctgaaggaataagacagtggtggaagctcaaacctggcccaccaccaccaaagcccgcagagc
ggcataaggacgacagcaggggtcttgtgcttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccg
gtcaacgaggcagacgccgcggccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacc
tcaagtacaaccacgccgacgcggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtct
tccaggcgaaaaagagggttcttgaacctctgggcctggttgaggaacctgttaagACCgctccgggaaaaaagaggccggt
agagcactctcctgtggagccagactcctcctcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaattttg
gtcagactggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctggtctCggaactaat
acCCTCgctacaggcagtggcgcaccaCTCgcagacaataacgagggcgccgacggagtgggtaattcctcgggaaat
tggcattgcgattccacatggCTCggcgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaac
cacctctacaaacaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagcacccttgggggtattttga
cttcaacagattccactgccacttttcaccacgtgactggcaaagactcatcaacaacaactggggattccgacccaagagactc
aacttcaagctctttaacattcaagtcaaagaggtcacgcagaatgacggtacgacgacgattgccaataaccttaccagcacg
gttcaggtgtttactgactcggagtaccagctcccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcaga
cgtcttcatggtgccacagtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagta
ctttccttctcagatgctgcgtaccggaaacaactttaccttcagctacacttttgaggacgttcctttccacagcagctacgctcaca
gccagagtctggaccgtctcatgaatcctctcatcgaccagtacctgtattacttgagcagaacaaacactccaagtggaaccac
cacgcagtcaaggcttcagttttctcaggccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttacc
gccagcagcgagtatcaaagacatctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaat
ggcagagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggggt
tctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaaaaggtcatgattacagacgaagaggaaatcagg
acaaccaatcccgtggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagcagctaccgcag
atgtcaacacacaaggcgttcttccaggcatggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagattcc
acacacggacggacattttcacccctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaaca
ccccggtacctgcgaatccttcgaccaccttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtcagc
gtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaatcccgaaattcagtacacttccaactacaacaagt
ctgttaatgtggactttactgtggacactaatggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaatctgtaa

Figure 20

CMV-VP2-delVP3:

gttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactt
acggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgcca
atagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagta
cgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagt
acatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgggg
atttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactcc
gccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccact
gcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcATGgctccgggaaaaaagaggccg
gtagagcactctcctgtggagccagactcctcctcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaattt
tggtcagactggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctggtctCggaacta
atacCCTCgctacaggcagtggcgcaccaCTCgcagacaataacgagggcgccgacggagtgggtaattcctcgggaa
attggcattgcgattccacatggCTCggcgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaa
ccacctctacaaacaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagcacccctgggggtattttg
acttcaacagattccactgccacttttcaccacgtgactggcaaagactcatcaacaacaactggggattccgacccaagagact
caacttcaagctctttaacattcaagtcaaagaggtcacgcagaatgacggtacgacgacgattgccaataaccttaccagcac
ggttcaggtgtttactgactcggagtaccagctcccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcag
acgtcttcatggtgccacagtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagt
actttccttctcagatgctgcgtaccggaaacaactttaccttcagctacacttttgaggacgttcctttccacagcagctacgctcac
agccagagtctggaccgtctcatgaatcctctcatcgaccagtacctgtattacttgagcagaacaaacactccaagtggaacca
ccacgcagtcaaggcttcagttttctcaggccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttac
cgccagcagcgagtatcaaagacatctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctca
atggcagagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggg
gttctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaaaaggtcatgattacagacgaagaggaaatcag
gacaaccaatcccgtggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagcagctaccgca
gatgtcaacacacaaggcgttcttccaggcatggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagatt
ccacacacggacggacattttcacccctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaa
caccccggtacctgcgaatccttcgaccaccttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtca
gcgtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaatcccgaaattcagtacacttccaactacaaca
agtctgttaatgtggactttactgtggacactaatggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaatctg
taattgcttgttaatcaataaaccgtttaattcgtttcagttgaactttggtctctgcgtatttctttcttatctagtttccatggctacgtagata
agtagcatggcgggttaatcattaactacagc

Figure 21

CMV-VP1-VP2-delVP3:

gttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataactt
acggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgcca
atagggactttccattgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagta
cgcccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagt
acatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacgggg
atttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactcc
gccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccact
gcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcATGgctgccgatggttatcttccagatt
ggctcgaggacactctctctgaaggaataagacagtggtggaagctcaaacctggcccaccaccaccaaagcccgcagagc
ggcataaggacgacagcaggggtcttgtgcttcctgggtacaagtacctcggacccttcaacggactcgacaagggagagccg
gtcaacgaggcagacgccgcggccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgtacc
tcaagtacaaccacgccgacgcggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtct
tccaggcgaaaaagagggttcttgaacctctgggcctggttgaggaacctgttaagACCgctccgggaaaaaagaggccggt
agagcactctcctgtggagccagactcctcctcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaattttg
gtcagactggagacgcagactcagtacctgacccccagcctctcggacagccaccagcagccccctctggtctCggaactaat
acCCTCgctacaggcagtggcgcaccaCTCgcagacaataacgagggcgccgacggagtgggtaattcctcgggaaat
tggcattgcgattccacatggCTCggcgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaac
cacctctacaaacaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagcacccttgggggtattttga
cttcaacagattccactgccacttttcaccacgtgactggcaaagactcatcaacaacaactggggattccgacccaagagactc
aacttcaagctctttaacattcaagtcaaagaggtcacgcagaatgacggtacgacgacgattgccaataaccttaccagcacg
gttcaggtgtttactgactcggagtaccagctcccgtacgtcctcggctcggcgcatcaaggatgcctcccgccgttcccagcaga
cgtcttcatggtgccacagtatggatacctcaccctgaacaacgggagtcaggcagtaggacgctcttcattttactgcctggagta
ctttccttctcagatgctgcgtaccggaaacaactttaccttcagctacacttttgaggacgttcctttccacagcagctacgctcaca
gccagagtctggaccgtctcatgaatcctctcatcgaccagtacctgtattacttgagcagaacaaacactccaagtggaaccac
cacgcagtcaaggcttcagttttctcaggccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttacc
gccagcagcgagtatcaaagacatctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaat
ggcagagactctctggtgaatccgggcccggccatggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggggt
tctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaaaaggtcatgattacagacgaagaggaaatcagg
acaaccaatcccgtggctacggagcagtatggttctgtatctaccaacctccagagaggcaacagacaagcagctaccgcag
atgtcaacacacaaggcgttcttccaggcatggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagattcc
acacacggacggacattttcacccctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaaca
ccccggtacctgcgaatccttcgaccaccttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtcagc
gtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaatcccgaaattcagtacacttccaactacaacaagt
ctgttaatgtggactttactgtggacactaatggcgtgtattcagagcctcgccccattggcaccagatacctgactcgtaatctgtaa
ttgcttgttaatcaataaaccgtttaattcgtttcagttgaactttggtctctgcgtatttctttcttatctagtttccatggctacgtagataagt
agcatggcgggttaatcattaactacaacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcat
agcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgt
caataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacggtaaactgcccactt
ggcagtacatcaagtgtatcatatgccaagtacgcccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagt
acatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatca
atgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatc
aacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataa
gcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctgg

Figure 22

ctagcATGgctccgggaaaaaagaggccggtagagcactctcctgtggagccagactcctcctcgggaaccggaaaggcg
ggccagcagcctgcaagaaaaagattgaattttggtcagactggagacgcagactcagtacctgacccccagcctctcggaca
gccaccagcagccccctctggtctCggaactaatacCCTCgctacaggcagtggcgcaccaCTCgcagacaataacga
gggcgccgacggagtgggtaattcctcgggaaattggcattgcgattccacatggCTCggcgacagagtcatcaccaccagc
acccgaacctgggccctgcccacctacaacaaccacctctacaaacaaatttccagccaatcaggagcctcgaacgacaatc
actactttggctacagcacccctttgggggtattttgacttcaacagattccactgccacttttcaccacgtgactggcaaagactcatc
aacaacaactggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaaagaggtcacgcagaatgacgg
tacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtaccagctcccgtacgtcctcggctcgg
cgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgccacagtatggatacctcaccctgaacaacgggagtc
aggcagtaggacgctcttcattttactgcctggagtactttccttctcagatgctgcgtaccggaaacaactttaccttcagctacactt
ttgaggacgttcctttccacagcagctacgctcacagccagagtctggaccgtctcatgaatcctctcatcgaccagtacctgtatta
cttgagcagaacaaacactccaagtggaaccaccacgcagtcaaggcttcagttttctcaggccggagcgagtgacattcggg
accagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaaagacatctgcggataacaacaacagtgaat
actcgtggactggagctaccaagtaccacctcaatggcagagactctctggtgaatccgggcccggccatggcaagccacaa
ggacgatgaagaaaagtttttttcctcagagcggggttctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaa
aaggtcatgattacagacgaagaggaaatcaggacaaccaatcccgtggctacggagcagtatggttctgtatctaccaacctc
cagagaggcaacagacaagcagctaccgcagatgtcaacacacaaggcgttcttccaggcatggtctggcaggacagagat
gtgtaccttcaggggcccatctgggcaaagattccacacacggacggacattttcacccctctcccctcatgggtggattcggactt
aaacaccctcctccacagattctcatcaagaacaccccggtacctgcgaatccttcgaccaccttcagtgcggcaaagtttgcttc
cttcatcacacagtactccacgggacaggtcagcgtggagatcgagtgggagctgcagaaggaaaacagcaaacgctggaa
tcccgaaattcagtacacttccaactacaacaagtctgttaatgtggactttactgtggacactaatggcgtgtattcagagcctcgc
cccattggcaccagatacctgactcgtaatctgtaattgcttgttaatcaataaaccgtttaattcgtttcagttgaactttggtctctgcg
tatttctttcttatctagtttccatggctacgtagataagtagcatggcgggttaatcattaactaca

Figure 22 continued

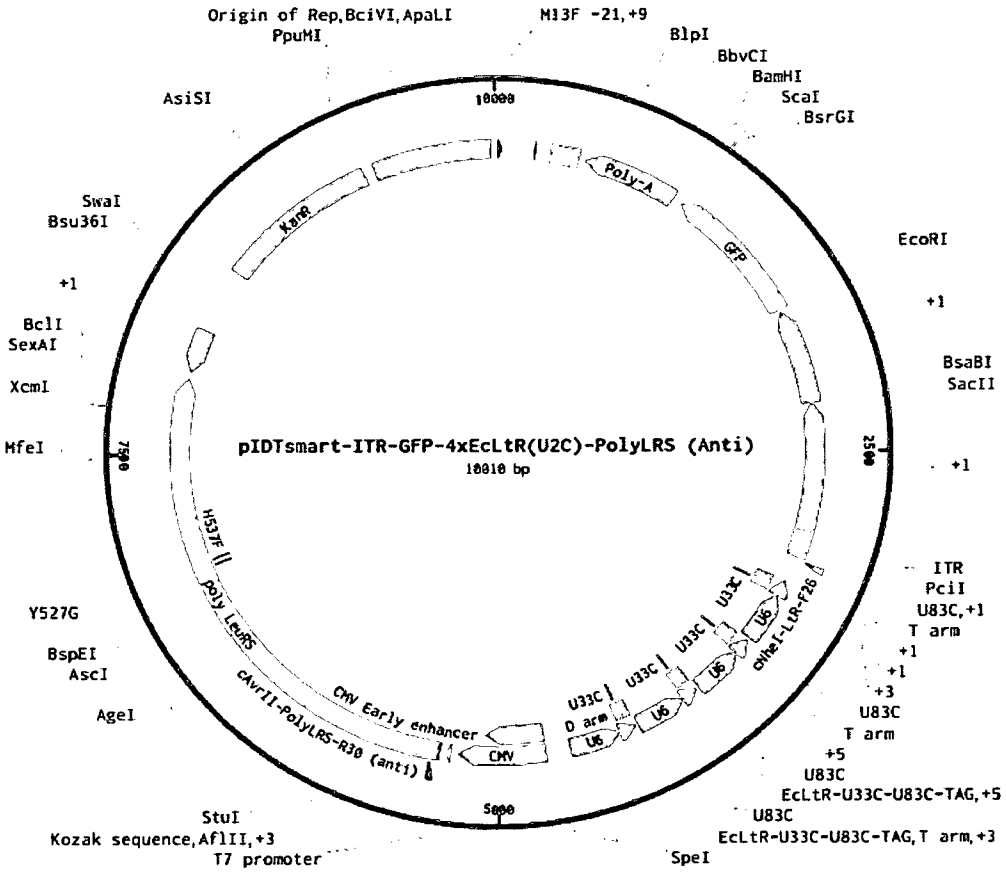


Figure 23A

pIDTsmart-ITR-GFP-4xEcLtR-LeuRS

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGACCG
GAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGTCCT
GAATGATATGCGaCCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTCGCG
CTGCTCTCGTCGATCCGctaggACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGC
ACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCCCTGCAGGCAGCTGCGCGCTC
GCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGC
CCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGG
GGTTCCTGCGGCCGCTCGGTCCGCACGTGGTTACCTACAAAATCAGAAGGACAGGGAA
GGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATC
ACCTGAGATTAGGAGTTGGAGACCAGCCTGGCCAATATGGTGAAACCCCGTCTCTACC
AAAAAAACAAAAATTAGCTGAGCCTGGTCATGCATGCCTGGAATCCCAACAACTCGGGA
GGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGATTGCAGTGAGCCAAGAT

Figure 23B

TGTGCCACTGCACTCCAGCTTGGTTCCCAATAGACCCCGCAGGCCCTACAGGTTGTCT
TCCCAACTTGCCCCTTGCTCCATACCACCCCCCTCCACCCCATAATATTATAGAAGGAC
ACCTAGTCAGACAAAATGATGCAACTTAATTTTATTAGGACAAGGCTGGTGGGCACTGG
AGTGGCAACTTCCAGGGCCAGGAGAGGCACTGGGGAGGGGTCACAGGGATGCCACC
CGTAGATCTCTCGAGCAGCGCTGCTCGAGGCAAGCTTCTGCAGGTCGACTCTAGAGGA
TCCCTACTTGAGCTCGAGATCTGAGTACTTGTACAGCTCGTCCATGCCGAGAGTGATCC
CGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTT
GCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGT
CGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGA
TGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATA
TAGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTC
CTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTC
ACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGA
CGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTA
GCGGCTGAAGCACTGCACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCA
CGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGTAGGTGGCAT
CGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGTCCAGCT
CGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATTCAGAA
TTCAATCGATGTTCGAATCCCAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACCAG
CACGTTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGCAAAAGG
GCCTAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATAAAAGCAGAATG
GTAGCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACATCAGTTACAATTTATA
TGCAGAAATATTTATATGCAGAAATATTGCTATTGCCTTAACCCAGAAATTATCACTGTTA
TTCTTTAGAATGGTGCAAAGAGGCATGATACATTGTATCATTATTGCCCTGAAAGAAAGA
GATTAGGGAAAGTATTAGAAATAAGATAAACAAAAAAGTATATTAAAAGAAGAAAGCATT
TTTTGTGGGCCTATAGACTCTATAGGCGGTACTTACGTCACTCTTGGCACGGGGAATCC
GCGTTCCAATGCACCGTTCCCGGCCGGGATTCGAATCCGCGGAGGCTGGATCGGTCC
CGGTGTCTTCTATGGAGGTCAAAACAGCGTGGATGGCGTCTCCAGGCGATCTGACGGT
TCACTAAACGAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGC
GTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGGTGCAAAAC
AAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCT
ATCCACGCCCATTGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTAAT
ACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCA
GGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACT
TGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAA
AGTCCCTATTGACGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGT
CGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCGGAACTCCAT
ATATGGGCTATGAACTAATGACCCCGTAATTGATTACTATTAATAACTAGCTCCACGCGT
GCGGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTC
GCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGG
CCTCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGACATGTGAGCAAAAGGCCAGC
AAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC
CCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGAgctagca
aaaaaTgCCCGGAGCGGGACTTGAACCCGCACAGCGCGAACGCCGAGGGATTtagAgTCC
CTTGTGTCTACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAA
AGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACA
TAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACT
AATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTA

Figure 23B continued

TATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAGGTACCcctagca
aaaaaTgCCCGGAGCGGGACTTGAACCCGCACAGCGCGAACGCCGAGGGATTtagAgTCC
CTTGTGTCTACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAA
AGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACA
TAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACT
AATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTA
TATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAGGTACCcctagca
aaaaaTgCCCGGAGCGGGACTTGAACCCGCACAGCGCGAACGCCGAGGGATTtagAgTCC
CTTGTGTCTACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAA
AGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACA
TAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACT
AATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTA
TATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAGGTACCcctagca
aaaaaTgCCCGGAGCGGGACTTGAACCCGCACAGCGCGAACGCCGAGGGATTtagAgTCC
CTTGTGTCTACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAA
AGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACA
TAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACT
AATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTA
TATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAGGTACCcCCTAG
Gtgcgctgcttcgcgatgtacgggccagatatacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcatta
gttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccatt
gacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggaGtatttacggtaaactgc
ccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatg
cccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagt
acatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgcaaatgggcggtaggcgtgtac
ggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcgaaattaatacgactcactata
gggagacccaagctggctagCGTTTAAACTTAAGCTTGCCGCCACCatggaagagcaataccgcccggaa
gagatagaatccaaagtacagcttcattgggatgagaagcgcacatttgaagtaaccgaagacgagagcaaagagaagtatt
actgcctgtctAtcCTCccctatccttctggtcgactacacatgggccacgtacgtaactacaccatcggtgacgtgatcgcccg
ctaccagcgtatgctgggcaaaaacgtcctgcagccgatcggctgggacgcgtttggtctgcctgcggaaggcgcggcggtga
aaaacaacaccgctccggcaccgtggacgtacgacaacatcgcgtatatgaaaaaccagctcaaaatgctgggctttggttatg
actggagccgcgagctggcaacctgtacgccggaatactaccgttgggaacagaaattcttcaccgagctgtataaaaaaggc
ctggtatataagaagacttctgcggtcaactggtgtccgaacgaccagaccgtactggcgaacgaacaagttatcgacggctgc
tgctggcgctgcgataccaaagttgaacgtaaagagatcccgcagtggtttatcaaaatcactgcttacgctgacgagctgctca
acgatctggataaactggatcactggccagacaccgttaaaaccatgcagcgtaactggatcggtcgttccgaaggcgtggag
atcaccttcaacgttaacgactatgacaacacgctgaccgtttacactacccgcccggacGCGtttatgggttgtacctacctgg
cggtagctgcgggtcatccgctggcgcagaaagcggcggaaaataatcctgaactggcggcctttattgacgaatgccgtaac
accaaagttgccgaagctgaaatggcgacgatggagaaaaaaggcgtcgatactggctttaaagcggttcacccattaacggg
cgaagaaattcccgtttgggcagcaaacttcgtattgatggagtacggcacgggcgcagttatggcggtaccggggcacgacc
agcgcgactacgagtttgcctctaaatacggcctgaacatcaaaccggttatcctggcagctgacggctctgagccagatctttct
cagcaagccctgactgaaaaaggcgtgctgttcaactctggcgagttcaacggtcttgaccatgaagcggccttcaacgccatc
gccgataaactgactgcgatgggcgttggcgagcgtaaagtgaactaccgcctgcgcgactggggtgtttcccgtcagcgttact
ggggcgcgccgattccgatggtgactctagaagacggtaccgtaatgccgaccccggacgaccagctgccggtgatcctgccg
gaagatgtggtaatggacggcattaccagcccgattaaagcagatccggagtgggcgaaaactaccgttaacggtatgccagc
actgcgtgaaaccgacactttcgacacctttatggagtcctcctggATCtatgcgcgctacacttgcccgcagtacaaagaaggt
atgctggattccgaagcggctaactactggctgccggtggatatcGCtattggtggtattgaacacgccattatgGGtctgctcta
cttccgcttcttccacaaactgatgcgtgatgcaggcatggtgaactctgacgaaccagcgaaacagttgctgtgtcagggtatgg
tgctggcagatgccttctactatgttggcgaaaacggcgaacgtaactgggtttccccggttgatgctatcgttgaacgtgacgaga

Figure 23B continued

aaggccgtatcgtgaaagcgaaagatgcggcaggccatgaactggtttataccggcatgagcaaaatgtccaagtcgaagaa
caacggtatcgacccgcaggtgatggttgaacgttacggcgcggacaccgttcgtctgtttatgatgtttgcttctccggctgatatg
actctcgaatggcaggaatccggtgtggaaggggctaaccgcttcctgaaacgtgtctggaaactggtttacgagcacacagca
aaaggtgatgttgcggcactgaacgttgatgcgctgactgaaaatcagaaagcgctgcgtcgcgatgtgcataaaacgatcgct
aaagtgaccgatgatatcggccgtcgtcagaccttcaacaccgcaattgcggcgattatggagctgatgaacaaactggcgaa
agcaccaaccgatggcgagcaggatcgcgctctgatgcaggaagcactgctggccgttgtccgtatgcttaacccgttcacccc
gcacatctgcttcacgctgtggcaggaactgaaaggcgaaggcgatatcgacaacgcgccgtggccggttgctgacgaaaaa
gcgatggtggaagactccacgctggtcgtggtgcaggttaacggtaaagtccgtgccaaaatcaccgttccggtggacgcaac
ggaagaacaggttcgcgaacgtgctggccaggaacatctggtagcaaaatatcttgatggcgttactgtacgtaaagtgatttac
gtaccaggtaaactcctcaatctggtcgttggcgggcccgtttaactcgagtctagagggcccgtttaaacccgctgatcagcctc
gactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcc
taataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcaggacagcaaggggg
aggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcggaaagaaccCCTAGGGG
TGCGAGCGGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACcCGT
GATCTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTT
GATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACCTCCAT
AGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGATCGGCACGTAAGAGGT
TCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGA
TTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAA
GCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATA
ATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGA
GTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCA
GGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACT
CCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACAGCATTCCAGGTATT
AGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCC
GGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTTCGTCTCG
CTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGA
GCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCT
CACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGG
GGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGAT
CTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTT
CAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGAT
GAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCGGGTGGGCCTTTC
TGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAT
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCC
TGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCC
GCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACT
TATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCG
GTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTT
GGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCC
GGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGC
GCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

Figure 23B continued

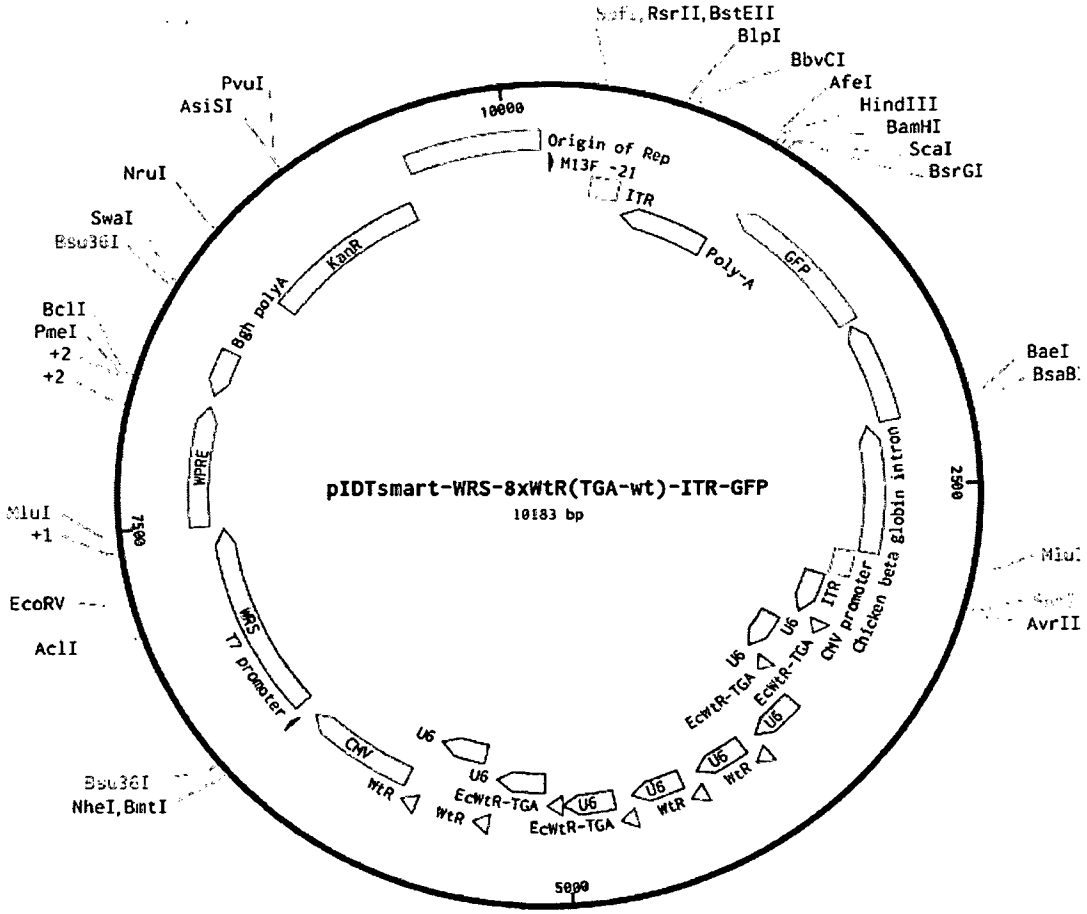


Figure 24A

pIDTsmart-TrpRS-8xWtR-ITR-GFP

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATCGTGGACCG
GAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGAGTCTGAGGCTCGTCCT
GAATGATATGCGaCCGCCGGAGGGTTGCGTTTGAGACGGGCGACAGATCCAGTCGCG
CTGCTCTCGTCGATCCGctaggCGGccGCaaataCCTGCAGGCAGCTGCGCGCTCGCTC
GCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGG
CCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTT
CCTGCGGCCGCTCGGTCCGCACGTGGTTACCTACAAAATCAGAAGGACAGGGAAGGG
AGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGGGAGGCCAAGGTGGGTAGATCACC
TGAGATTAGGAGTTGGAGACCAGCCTGGCCAATATGGTGAAACCCCGTCTCTACCAAA
AAAACAAAAATTAGCTGAGCCTGGTCATGCATGCCTGGAATCCCAACAACTCGGGAGG
CTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGATTGCAGTGAGCCAAGATTG

Figure 24B

TGCCACTGCACTCCAGCTTGGTTCCCAATAGACCCCGCAGGCCCTACAGGTTGTCTTC
CCAACTTGCCCCTTGCTCCATACCACCCCCCTCCACCCCATAATATTATAGAAGGACAC
CTAGTCAGACAAAATGATGCAACTTAATTTTATTAGGACAAGGCTGGTGGGCACTGGAG
TGGCAACTTCCAGGGCCAGGAGAGGCACTGGGGAGGGGTCACAGGGATGCCACCCGT
AGATCTCTCGAGCAGCGCTGCTCGAGGCAAGCTTCTGCAGGTCGACTCTAGAGGATCC
CTACTTGAGCTCGAGATCTGAGTACTTGTACAGCTCGTCCATGCCGAGAGTGATCCCG
GCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGC
TCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCG
CCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGATG
TTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATA
GACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCT
TGAAGTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCAC
CTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACG
TAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGC
GGCTGAAGCACTGCACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCACG
GGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGTAGGTGGCATCG
CCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGTCCAGCTCGA
CCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATTCAGAATTC
AATCGATGTTCGAATCCCAATTCTTTGCCAAAGTGATGGGCCAGCACACAGACCAGCAC
GTTGCCCAGGAGCTGTGGGAGGAAGATAAGAGGTATGAACATGATTAGCAAAAGGGCC
TAGCTTGGACTCAGAATAATCCAGCCTTATCCCAACCATAAAATAAAAGCAGAATGGTA
GCTGGATTGTAGCTGCTATTAGCAATATGAAACCTCTTACATCAGTTACAATTTATATGC
AGAAATATTTATATGCAGAAATATTGCTATTGCCTTAACCCAGAAATTATCACTGTTATTC
TTTAGAATGGTGCAAAGAGGCATGATACATTGTATCATTATTGCCCTGAAAGAAAGAGAT
TAGGGAAAGTATTAGAAATAAGATAAACAAAAAAGTATATTAAAAGAAGAAAGCATTTTTT
GTGGGCCTATAGACTCTATAGGCGGTACTTACGTCACTCTTGGCACGGGGAATCCGCG
TTCCAATGCACCGTTCCCGGCCGGGATTCGAATCCGCGGAGGCTGGATCGGTCCCGG
TGTCTTCTATGGAGGTCAAAACAGCGTGGATGGCGTCTCCAGGCGATCTGACGGTTCA
CTAAACGAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTC
AATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGGTGCAAAACAAA
CTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATC
CACGCCCATTGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTAATACG
TAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATAATGCCAGG
CGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGA
TGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGT
CCCTATTGACGTTACTATGGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGT
TGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCGGAACTCCATATA
TGGGCTATGAACTAATGACCCCGTAATTGATTACTATTAATAACTAGCTCCACGCGTGC
GGCCGCAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGC
TCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCC
TCAGTGAGCGAGCGAGCGCGCAGCTGCCTGCAGGCTCGATCCGCTCGCACCCtagGTC
GGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC
TGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATAC

Figure 24B continued

GTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATG
GACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGT
GGAAAGGACGAAACACCAGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTtcAAAACCG
GGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAttttttgCTAGGTCGGGCAGGAAGAGG
GCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAA
TTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGT
AATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCT
TACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAA
CACCAGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTtcAAAACCGGGTGtTGGGAGTT
CGAGTCTCTCCGCCCCTGCCAttttttgCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCAT
GATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTT
GACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG
GTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGA
AAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCAGGGGCGT
AGTTCAATTGGTAGAGCACCGGTCTtcAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGC
CCCTGCCAttttttgCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTT
GCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAA
GATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTT
AAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTC
TTGGCTTTATATATCTTGTGGAAAGGACGAAACACCAGGGGCGTAGTTCAATTGGTAGA
GCACCGGTCTtcAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAttttttgctag
GTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAA
GGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAA
TACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAA
ATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCT
TGTGGAAAGGACGAAACACCAGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTtcAAAA
CCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAttttttgCTAGGTCGGGCAGGAAG
AGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGA
TAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAA
AGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATAT
GCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
AAACACCAGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTtcAAAACCGGGTGtTGGGA
GTTCGAGTCTCTCCGCCCCTGCCAttttttgCTAGGTCGGGCAGGAAGAGGGCCTATTTCC
CATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAA
TTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTT
GGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTT
GAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCAGGGGC
GTAGTTCAATTGGTAGAGCACCGGTCTtcAAAACCGGGTGtTGGGAGTTCGAGTCTCTCC
GCCCCTGCCAttttttgCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATA
TTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACAC
AAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGT
TTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGAT
TTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCAGGGGCGTAGTTCAATTGGT
AGAGCACCGGTCTtcAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAttttttg
ctAGGttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcc

Figure 24B continued

cgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttcca
ttgacgtcaatgggtggactatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgac
gtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagt
catcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccac
cccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgca
aatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactagagaacccactgcttactggcttatcg
aaattaatacgactcactatagggagacccaagctggctagcgccaccATGACTAAGCCCATCGTTTTTgctGG
CGCACAGCCCTCAGGTGAATTGACCATTGGTAACTACATGGGTGCGCTGCGTCAGTGG
GTAAACATGCAGGATGACTACCATTGCATTTACTGTATCGTTGACCAACACGCGATCAC
CGTGCGCCAGGATGCACAGAAGCTGCGTAAAGCGACGCTGGATACGCTGGCCTTGTAT
CTGGCTTGTGGTATCGATCCTGAGAAAAGCACCATTTTTGTTCAGTCCCACGTGCCGGA
ACATGCACAGTTAGGCTGGGCACTGAACTGCTATACCTACTTCGGCGAACTGAGTCGC
ATGACGCAGTTTAAAGATAAATCTGCGCGTTATGCCGAGAACATCAACGCTGGTCTGTT
TGACTATCCGGTGCTGATGGCAGCGGACATCCTGCTGTATCAAACTAATCTGggtccttgtG
GTGAAGACCAGAAACAGCACCTCGAACTGAGCCGCGATATTGCCCAGCGTTTCAACGC
GCTGTATGGCGAGATCTTTAAGGTGCCGGAGCCGTTTATTCCGAAATCTGGCGCGCGC
GTAATGTCGCTGCTGGAGCCGACCAAGAAGATGTCCAAGTCTGACGATAATCGCAATA
ACGTTATCGGCCTGCTGGAAGATCCGAAATCGGTAGTGAAGAAAATCAAACGTGCGGT
CACTGACTCCGACGAGCCGCCGGTAGTTCGCTACGATGTGCAGAACAAAGCGGGCGT
TTCCAACCTGTTGGATATCCTTTCAGCGGTAACGGGCCAGAGCATCCCAGAACTGGAA
AAACAGTTCGAAGGCAAGATGTATGGTCATCTGAAAGGTGAAGTGGCTGATGCCGTTTC
CGGTATGCTGACTGAATTGCAGGAACGCTATCACCGTTTCCGCAACGATGAAGCCTTC
CTGCAACAGGTGATGAAAGATGGCGCGGAAAAAGCCAGCGCGCACGCTTCCCGTACG
CTAAAAGCGGTGTACGAAGCGATTGGTTTTGTGGCGAAGCCGTAAgaattcaacgcgttaagtcg
acaatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgcttta
atgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggccc
gttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctccttt
ccgggactttcgctttccccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgt
tgggcactgacaattccgtggtgttgtcggggaaatcatcgtcctttccttggctgctcgcctgtgttgccacctggattctgcgcggg
acgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtct
tcgccttcgccctcagacgagtcggatctccctttgggccgcctccccgcgtcgactttaactcgagtctagagggcccgtttaaac
ccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgcca
ctcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggcag
gacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggaCTAGGGGTGCG
AGCGGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACcCGTGATC
TTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTTGATGT
CATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACCTCCATAGCAG
AAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGATCGGCACGTAAGAGGTTCCAA
CTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTC
AGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCG
CGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTC
GGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGT
TTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGGCTA
AACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGAT
GATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACAGCATTCCAGGTATTAGAAGA

Figure 24B continued

ATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGC
ATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTTCGTCTCGCTCAGG
CGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAAT
GGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGA
TTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATT
AATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCA
TCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAAT
ATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGATGAGTTTTT
CTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTGC
TGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGATGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT
CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTG
TAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCT
GCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCA
CTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACA
GAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTG
CGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAC
AAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAA
AAAGGATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

Figure 24B continued

SEQUENCE LISTING

<110> TRUSTEES OF BOSTON COLLEGE

<120> CONTROLLED MODIFICATION OF ADENO-ASSOCIATED VIRUS (AAV) FOR ENHANCED GENE THERAPY

<130> 0342.0010WO1

<140>
<141>

<150> 63/018,573
<151> 2020-05-01

<160> 11

<170> PatentIn version 3.5

<210> 1
<211> 735
<212> PRT
<213> Adeno-associated virus 2

<400> 1
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Thr Leu Ser
1 5 10 15

Glu Gly Ile Arg Gln Trp Trp Lys Leu Lys Pro Gly Pro Pro Pro Pro
            20 25 30

Lys Pro Ala Glu Arg His Lys Asp Asp Ser Arg Gly Leu Val Leu Pro
        35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Phe Asn Gly Leu Asp Lys Gly Glu Pro
    50 55 60

Val Asn Glu Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Arg Gln Leu Asp Ser Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
                85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
            100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Val Leu Glu Pro
        115 120 125

Leu Gly Leu Val Glu Glu Pro Val Lys Thr Ala Pro Gly Lys Lys Arg
    130 135 140

Pro Val Glu His Ser Pro Val Glu Pro Asp Ser Ser Ser Gly Thr Gly
145 150 155 160

Lys Ala Gly Gln Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln Thr
                165 170 175

```
Gly Asp Ala Asp Ser Val Pro Asp Pro Gln Pro Leu Gly Gln Pro Pro
            180                 185                 190

Ala Ala Pro Ser Gly Leu Gly Thr Asn Thr Met Ala Thr Gly Ser Gly
        195                 200                 205

Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Asn Ser
    210                 215                 220

Ser Gly Asn Trp His Cys Asp Ser Thr Trp Met Gly Asp Arg Val Ile
225                 230                 235                 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
                245                 250                 255

Tyr Lys Gln Ile Ser Ser Gln Ser Gly Ala Ser Asn Asp Asn His Tyr
            260                 265                 270

Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg Phe His
        275                 280                 285

Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn Asn Trp
    290                 295                 300

Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile Gln Val
305                 310                 315                 320

Lys Glu Val Thr Gln Asn Asp Gly Thr Thr Thr Ile Ala Asn Asn Leu
                325                 330                 335

Thr Ser Thr Val Gln Val Phe Thr Asp Ser Glu Tyr Gln Leu Pro Tyr
            340                 345                 350

Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe Pro Ala Asp
        355                 360                 365

Val Phe Met Val Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asn Gly Ser
    370                 375                 380

Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe Pro Ser
385                 390                 395                 400

Gln Met Leu Arg Thr Gly Asn Asn Phe Thr Phe Ser Tyr Thr Phe Glu
                405                 410                 415

Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu Asp Arg
            420                 425                 430

Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser Arg Thr
        435                 440                 445
```

```
Asn Thr Pro Ser Gly Thr Thr Thr Gln Ser Arg Leu Gln Phe Ser Gln
    450                 455                 460

Ala Gly Ala Ser Asp Ile Arg Asp Gln Ser Arg Asn Trp Leu Pro Gly
465                 470                 475                 480

Pro Cys Tyr Arg Gln Gln Arg Val Ser Lys Thr Ser Ala Asp Asn Asn
                485                 490                 495

Asn Ser Glu Tyr Ser Trp Thr Gly Ala Thr Lys Tyr His Leu Asn Gly
            500                 505                 510

Arg Asp Ser Leu Val Asn Pro Gly Pro Ala Met Ala Ser His Lys Asp
        515                 520                 525

Asp Glu Glu Lys Phe Phe Pro Gln Ser Gly Val Leu Ile Phe Gly Lys
    530                 535                 540

Gln Gly Ser Glu Lys Thr Asn Val Asp Ile Glu Lys Val Met Ile Thr
545                 550                 555                 560

Asp Glu Glu Glu Ile Arg Thr Thr Asn Pro Val Ala Thr Glu Gln Tyr
                565                 570                 575

Gly Ser Val Ser Thr Asn Leu Gln Arg Gly Asn Arg Gln Ala Ala Thr
            580                 585                 590

Ala Asp Val Asn Thr Gln Gly Val Leu Pro Gly Met Val Trp Gln Asp
        595                 600                 605

Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His Thr
    610                 615                 620

Asp Gly His Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Leu Lys
625                 630                 635                 640

His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala Asn
                645                 650                 655

Pro Ser Thr Thr Phe Ser Ala Ala Lys Phe Ala Ser Phe Ile Thr Gln
            660                 665                 670

Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln Lys
        675                 680                 685

Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn Tyr
    690                 695                 700

Asn Lys Ser Val Asn Val Asp Phe Thr Val Asp Thr Asn Gly Val Tyr
705                 710                 715                 720

Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
```

725 730 735

<210> 2
<211> 598
<212> PRT
<213> Adeno-associated virus 2

<400> 2
Met Ala Pro Gly Lys Lys Arg Pro Val Glu His Ser Pro Val Glu Pro
1 5 10 15

Asp Ser Ser Ser Gly Thr Gly Lys Ala Gly Gln Gln Pro Ala Arg Lys
20 25 30

Arg Leu Asn Phe Gly Gln Thr Gly Asp Ala Asp Ser Val Pro Asp Pro
35 40 45

Gln Pro Leu Gly Gln Pro Pro Ala Ala Pro Ser Gly Leu Gly Thr Asn
50 55 60

Thr Met Ala Thr Gly Ser Gly Ala Pro Met Ala Asp Asn Asn Glu Gly
65 70 75 80

Ala Asp Gly Val Gly Asn Ser Ser Gly Asn Trp His Cys Asp Ser Thr
85 90 95

Trp Met Gly Asp Arg Val Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu
100 105 110

Pro Thr Tyr Asn Asn His Leu Tyr Lys Gln Ile Ser Ser Gln Ser Gly
115 120 125

Ala Ser Asn Asp Asn His Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr
130 135 140

Phe Asp Phe Asn Arg Phe His Cys His Phe Ser Pro Arg Asp Trp Gln
145 150 155 160

Arg Leu Ile Asn Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe
165 170 175

Lys Leu Phe Asn Ile Gln Val Lys Glu Val Thr Gln Asn Asp Gly Thr
180 185 190

Thr Thr Ile Ala Asn Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp
195 200 205

Ser Glu Tyr Gln Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly Cys
210 215 220

Leu Pro Pro Phe Pro Ala Asp Val Phe Met Val Pro Gln Tyr Gly Tyr
225 230 235 240

```
Leu Thr Leu Asn Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr
                245                 250                 255

Cys Leu Glu Tyr Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe
            260                 265                 270

Thr Phe Ser Tyr Thr Phe Glu Asp Val Pro Phe His Ser Ser Tyr Ala
        275                 280                 285

His Ser Gln Ser Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr
    290                 295                 300

Leu Tyr Tyr Leu Ser Arg Thr Asn Thr Pro Ser Gly Thr Thr Thr Gln
305                 310                 315                 320

Ser Arg Leu Gln Phe Ser Gln Ala Gly Ala Ser Asp Ile Arg Asp Gln
                325                 330                 335

Ser Arg Asn Trp Leu Pro Gly Pro Cys Tyr Arg Gln Gln Arg Val Ser
            340                 345                 350

Lys Thr Ser Ala Asp Asn Asn Asn Ser Glu Tyr Ser Trp Thr Gly Ala
        355                 360                 365

Thr Lys Tyr His Leu Asn Gly Arg Asp Ser Leu Val Asn Pro Gly Pro
    370                 375                 380

Ala Met Ala Ser His Lys Asp Asp Glu Glu Lys Phe Phe Pro Gln Ser
385                 390                 395                 400

Gly Val Leu Ile Phe Gly Lys Gln Gly Ser Glu Lys Thr Asn Val Asp
                405                 410                 415

Ile Glu Lys Val Met Ile Thr Asp Glu Glu Glu Ile Arg Thr Thr Asn
            420                 425                 430

Pro Val Ala Thr Glu Gln Tyr Gly Ser Val Ser Thr Asn Leu Gln Arg
        435                 440                 445

Gly Asn Arg Gln Ala Ala Thr Ala Asp Val Asn Thr Gln Gly Val Leu
    450                 455                 460

Pro Gly Met Val Trp Gln Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile
465                 470                 475                 480

Trp Ala Lys Ile Pro His Thr Asp Gly His Phe His Pro Ser Pro Leu
                485                 490                 495

Met Gly Gly Phe Gly Leu Lys His Pro Pro Pro Gln Ile Leu Ile Lys
            500                 505                 510

Asn Thr Pro Val Pro Ala Asn Pro Ser Thr Thr Phe Ser Ala Ala Lys
```

```
        515                 520                 525

Phe Ala Ser Phe Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu
    530                 535                 540

Ile Glu Trp Glu Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu
545                 550                 555                 560

Ile Gln Tyr Thr Ser Asn Tyr Asn Lys Ser Val Asn Val Asp Phe Thr
                565                 570                 575

Val Asp Thr Asn Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg
            580                 585                 590

Tyr Leu Thr Arg Asn Leu
        595


<210> 3
<211> 533
<212> PRT
<213> Adeno-associated virus 2

<400> 3
Met Ala Thr Gly Ser Gly Ala Pro Met Ala Asp Asn Asn Glu Gly Ala
1               5                   10                  15

Asp Gly Val Gly Asn Ser Ser Gly Asn Trp His Cys Asp Ser Thr Trp
            20                  25                  30

Met Gly Asp Arg Val Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro
        35                  40                  45

Thr Tyr Asn Asn His Leu Tyr Lys Gln Ile Ser Ser Gln Ser Gly Ala
    50                  55                  60

Ser Asn Asp Asn His Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe
65                  70                  75                  80

Asp Phe Asn Arg Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg
                85                  90                  95

Leu Ile Asn Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys
            100                 105                 110

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

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

Glu Tyr Gln Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly Cys Leu
145                 150                 155                 160
```

```
Pro Pro Phe Pro Ala Asp Val Phe Met Val Pro Gln Tyr Gly Tyr Leu
                165                 170                 175

Thr Leu Asn Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys
            180                 185                 190

Leu Glu Tyr Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Thr
        195                 200                 205

Phe Ser Tyr Thr Phe Glu Asp Val Pro Phe His Ser Ser Tyr Ala His
    210                 215                 220

Ser Gln Ser Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu
225                 230                 235                 240

Tyr Tyr Leu Ser Arg Thr Asn Thr Pro Ser Gly Thr Thr Thr Gln Ser
                245                 250                 255

Arg Leu Gln Phe Ser Gln Ala Gly Ala Ser Asp Ile Arg Asp Gln Ser
            260                 265                 270

Arg Asn Trp Leu Pro Gly Pro Cys Tyr Arg Gln Gln Arg Val Ser Lys
        275                 280                 285

Thr Ser Ala Asp Asn Asn Asn Ser Glu Tyr Ser Trp Thr Gly Ala Thr
    290                 295                 300

Lys Tyr His Leu Asn Gly Arg Asp Ser Leu Val Asn Pro Gly Pro Ala
305                 310                 315                 320

Met Ala Ser His Lys Asp Asp Glu Glu Lys Phe Phe Pro Gln Ser Gly
                325                 330                 335

Val Leu Ile Phe Gly Lys Gln Gly Ser Glu Lys Thr Asn Val Asp Ile
            340                 345                 350

Glu Lys Val Met Ile Thr Asp Glu Glu Glu Ile Arg Thr Thr Asn Pro
        355                 360                 365

Val Ala Thr Glu Gln Tyr Gly Ser Val Ser Thr Asn Leu Gln Arg Gly
    370                 375                 380

Asn Arg Gln Ala Ala Thr Ala Asp Val Asn Thr Gln Gly Val Leu Pro
385                 390                 395                 400

Gly Met Val Trp Gln Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp
                405                 410                 415

Ala Lys Ile Pro His Thr Asp Gly His Phe His Pro Ser Pro Leu Met
            420                 425                 430

Gly Gly Phe Gly Leu Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn
```

```
        435                 440                 445

Thr Pro Val Pro Ala Asn Pro Ser Thr Thr Phe Ser Ala Ala Lys Phe
    450                 455                 460

Ala Ser Phe Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile
465                 470                 475                 480

Glu Trp Glu Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile
                485                 490                 495

Gln Tyr Thr Ser Asn Tyr Asn Lys Ser Val Asn Val Asp Phe Thr Val
            500                 505                 510

Asp Thr Asn Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr
        515                 520                 525

Leu Thr Arg Asn Leu
    530
```

<210> 4
<211> 4642
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic polynucleotide"

<400> 4

```
gctctagaac tagtggatcc cccggaagat cagaagttcc tattccgaag ttcctattct       60

ctagaaagta taggaacttc tgatctgcgc agccgccatg ccggggtttt acgagattgt      120

gattaaggtc cccagcgacc ttgacgagca tctgcccggc atttctgaca gctttgtgaa      180

ctgggtggcc gagaaggaat gggagttgcc gccagattct gacatggatc tgaatctgat      240

tgagcaggca cccctgaccg tggccgagaa gctgcagcgc gactttctga cggaatggcg      300

ccgtgtgagt aaggccccgg aggccctttt ctttgtgcaa tttgagaagg gagagagcta      360

cttccacatg cacgtgctcg tggaaaccac cggggtgaaa tccatggttt tgggacgttt      420

cctgagtcag attcgcgaaa aactgattca gagaatttac cgcgggatcg agccgacttt      480

gccaaactgg ttcgcggtca caaagaccag aaatggcgcc ggaggcggga acaaggtggt      540

ggatgagtgc tacatcccca attacttgct ccccaaaacc cagcctgagc tccagtgggc      600

gtggactaat atggaacagt atttaagcgc ctgtttgaat ctcacggagc gtaaacggtt      660

ggtggcgcag catctgacgc acgtgtcgca gacgcaggag cagaacaaag agaatcagaa      720

tcccaattct gatgcgccgg tgatcagatc aaaaacttca gccaggtaca tggagctggt      780

cgggtggctc gtggacaagg ggattacctc ggagaagcag tggatccagg aggaccaggc      840

ctcatacatc tccttcaatg cggcctccaa ctcgcggtcc caaatcaagg ctgccttgga      900

caatgcggga aagattatga gcctgactaa aaccgccccc gactacctgg tgggccagca      960
```

gcccgtggag gacatttcca gcaatcggat ttataaaatt ttggaactaa acgggtacga 1020

tccccaatat gcggcttccg tctttctggg atgggccacg aaaaagttcg gcaagaggaa 1080

caccatctgg ctgtttgggc ctgcaactac cgggaagacc aacatcgcgg aggccatagc 1140

ccacactgtg cccttctacg ggtgcgtaaa ctggaccaat gagaactttc ccttcaacga 1200

ctgtgtcgac aagatggtga tctggtggga ggaggggaag atgaccgcca aggtcgtgga 1260

gtcggccaaa gccattctcg gaggaagcaa ggtgcgcgtg gaccagaaat gcaagtcctc 1320

ggcccagata gacccgactc ccgtgatcgt cacctccaac accaacatgt gcgccgtgat 1380

tgacgggaac tcaacgacct tcgaacacca gcagccgttg caagaccgga tgttcaaatt 1440

tgaactcacc cgccgtctgg atcatgactt tgggaaggtc accaagcagg aagtcaaaga 1500

ctttttccgg tgggcaaagg atcacgtggt tgaggtggag catgaattct acgtcaaaaa 1560

gggtggagcc aagaaaagac ccgcccccag tgacgcagat ataagtgagc ccaaacgggt 1620

gcgcgagtca gttgcgcagc catcgacgtc agacgcggaa gcttcgatca actacgcaga 1680

caggtaccaa aacaaatgtt ctcgtcacgt gggcatgaat ctgatgctgt ttccctgcag 1740

acaatgcgag agaatgaatc agaattcaaa tatctgcttc actcacggac agaaagactg 1800

tttagagtgc tttcccgtgt cagaatctca acccgtttct gtcgtcaaaa aggcgtatca 1860

gaaactgtgc tacattcatc atatcatggg aaaggtgcca gacgcttgca ctgcctgcga 1920

tctggtcaat gtggatttgg atgactgcat ctttgaacaa taaatgattt aaatcaggtc 1980

tcgctgccga tggttatctt ccagattggc tcgaggacac tctctctgaa ggaataagac 2040

agtggtggaa gctcaaacct ggcccaccac caccaaagcc cgcagagcgg cataaggacg 2100

acagcagggg tcttgtgctt cctgggtaca agtacctcgg acccttcaac ggactcgaca 2160

agggagagcc ggtcaacgag gcagacgccg cggccctcga gcacgacaaa gcctacgacc 2220

ggcagctcga cagcggagac aacccgtacc tcaagtacaa ccacgccgac gcggagtttc 2280

aggagcgcct taaagaagat acgtcttttg ggggcaacct cggacgagca gtcttccagg 2340

cgaaaaagag ggttcttgaa cctctgggcc tggttgagga acctgttaag acggctccgg 2400

gaaaaaagag gccggtagag cactctcctg tggagccaga ctcctcctcg ggaaccggaa 2460

aggcgggcca gcagcctgca agaaaaagat tgaattttgg tcagactgga gacgcagact 2520

cagtacctga cccccagcct ctcggacagc caccagcagc cccctctggt ctgggaacta 2580

atacgatggc tacaggcagt ggcgcaccaa tggcagacaa taacgagggc gccgacggag 2640

tgggtaattc ctcgggaaat tggcattgcg attccacatg gatgggcgac agagtcatca 2700

ccaccagcac ccgaacctgg gccctgccca cctacaacaa ccacctctac aaacaaattt 2760

ccagccaatc aggagcctcg aacgacaatc actactttgg ctacagcacc ccttgggggt 2820

attttgactt caacagattc cactgccact tttcaccacg tgactggcaa agactcatca 2880

acaacaactg gggattccga cccaagagac tcaacttcaa gctctttaac attcaagtca 2940

aagaggtcac gcagaatgac ggtacgacga cgattgccaa taaccttacc agcacggttc 3000

aggtgtttac tgactcggag taccagctcc cgtacgtcct cggctcggcg catcaaggat 3060

gcctcccgcc gttcccagca gacgtcttca tggtgccaca gtatggatac ctcaccctga 3120

acaacgggag tcaggcagta ggacgctctt cattttactg cctggagtac tttccttctc 3180

agatgctgcg taccggaaac aactttacct tcagctacac ttttgaggac gttcctttcc 3240

acagcagcta cgctcacagc cagagtctgg accgtctcat gaatcctctc atcgaccagt 3300

acctgtatta cttgagcaga acaaacactc caagtggaac caccacgcag tcaaggcttc 3360

agttttctca ggccggagcg agtgacattc gggaccagtc taggaactgg cttcctggac 3420

cctgttaccg ccagcagcga gtatcaaaga catctgcgga taacaacaac agtgaatact 3480

cgtggactgg agctaccaag taccacctca atggcagaga ctctctggtg aatccgggcc 3540

cggccatggc aagccacaag gacgatgaag aaaagttttt tcctcagagc ggggttctca 3600

tctttgggaa gcaaggctca gagaaaacaa atgtggacat tgaaaaggtc atgattacag 3660

acgaagagga aatcaggaca accaatcccg tggctacgga gcagtatggt tctgtatcta 3720

ccaacctcca gagaggcaac agacaagcag ctaccgcaga tgtcaacaca caaggcgttc 3780

ttccaggcat ggtctggcag gacagagatg tgtaccttca ggggcccatc tgggcaaaga 3840

ttccacacac ggacggacat tttcacccct ctcccctcat gggtggattc ggacttaaac 3900

accctcctcc acagattctc atcaagaaca ccccggtacc tgcgaatcct tcgaccacct 3960

tcagtgcggc aaagtttgct tccttcatca cacagtactc cacgggacag gtcagcgtgg 4020

agatcgagtg ggagctgcag aaggaaaaca gcaaacgctg gaatcccgaa attcagtaca 4080

cttccaacta caacaagtct gttaatgtgg actttactgt ggacactaat ggcgtgtatt 4140

cagagcctcg ccccattggc accagatacc tgactcgtaa tctgtaattg cttgttaatc 4200

aataaaccgt ttaattcgtt tcagttgaac tttggtctct gcgtatttct ttcttatcta 4260

gtttccatgg ctacgtagat aagtagcatg gcgggttaat cattaactac agcccgggcg 4320

tttaaacagc gggcggaggg gtggagtcgt gacgtgaatt acgtcatagg gttagggagg 4380

tcctgtatta gaggtcacgt gagtgttttg cgacattttg cgacaccatg tggtctcgct 4440

gggggggggg gcccgagtga gcacgcaggg tctccatttt gaagcgggag gtttgaacga 4500

gcgctggcgc gctcactggc cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt 4560

acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa tagcgaagag 4620

gcccgcaccg atcgcccttc cc 4642

<210> 5
<211> 4642
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 5
gctctagaac tagtggatcc cccggaagat cagaagttcc tattccgaag ttcctattct 60

```
ctagaaagta taggaacttc tgatctgcgc agccgccatg ccggggtttt acgagattgt      120
gattaaggtc cccagcgacc ttgacgagca tctgcccggc atttctgaca gctttgtgaa      180
ctgggtggcc gagaaggaat gggagttgcc gccagattct gacatggatc tgaatctgat      240
tgagcaggca cccctgaccg tggccgagaa gctgcagcgc gactttctga cggaatggcg      300
ccgtgtgagt aaggccccgg aggccctttt ctttgtgcaa tttgagaagg gagagagcta      360
cttccacatg cacgtgctcg tggaaaccac cggggtgaaa tccatggttt tgggacgttt      420
cctgagtcag attcgcgaaa aactgattca gagaatttac cgcgggatcg agccgacttt      480
gccaaactgg ttcgcggtca caaagaccag aaatggcgcc ggaggcggga acaaggtggt      540
ggatgagtgc tacatcccca attacttgct ccccaaaacc cagcctgagc tccagtgggc      600
gtggactaat atggaacagt atttaagcgc ctgtttgaat ctcacggagc gtaaacggtt      660
ggtggcgcag catctgacgc acgtgtcgca gacgcaggag cagaacaaag agaatcagaa      720
tcccaattct gatgcgccgg tgatcagatc aaaaacttca gccaggtaca tggagctggt      780
cgggtggctc gtggacaagg ggattacctc ggagaagcag tggatccagg aggaccaggc      840
ctcatacatc tccttcaatg cggcctccaa ctcgcggtcc caaatcaagg ctgccttgga      900
caatgcggga aagattatga gcctgactaa aaccgccccc gactacctgg tgggccagca      960
gcccgtggag gacatttcca gcaatcggat ttataaaatt ttggaactaa acgggtacga     1020
tccccaatat gcggcttccg tctttctggg atgggccacg aaaaagttcg gcaagaggaa     1080
caccatctgg ctgtttgggc ctgcaactac cgggaagacc aacatcgcgg aggccatagc     1140
ccacactgtg cccttctacg ggtgcgtaaa ctggaccaat gagaactttc ccttcaacga     1200
ctgtgtcgac aagatggtga tctggtggga ggaggggaag atgaccgcca aggtcgtgga     1260
gtcggccaaa gccattctcg gaggaagcaa ggtgcgcgtg gaccagaaat gcaagtcctc     1320
ggcccagata gacccgactc ccgtgatcgt cacctccaac accaacatgt gcgccgtgat     1380
tgacgggaac tcaacgacct tcgaacacca gcagccgttg caagaccgga tgttcaaatt     1440
tgaactcacc cgccgtctgg atcatgactt tgggaaggtc accaagcagg aagtcaaaga     1500
ctttttccgg tgggcaaagg atcacgtggt tgaggtggag catgaattct acgtcaaaaa     1560
gggtggagcc aagaaaagac ccgcccccag tgacgcagat ataagtgagc ccaaacgggt     1620
gcgcgagtca gttgcgcagc catcgacgtc agacgcggaa gcttcgatca actacgcaga     1680
caggtaccaa aacaaatgtt ctcgtcacgt gggcatgaat ctgatgctgt ttccctgcag     1740
acaatgcgag agaatgaatc agaattcaaa tatctgcttc actcacggac agaaagactg     1800
tttagagtgc tttcccgtgt cagaatctca acccgtttct gtcgtcaaaa aggcgtatca     1860
gaaactgtgc tacattcatc atatcatggg aaaggtgcca gacgcttgca ctgcctgcga     1920
tctggtcaat gtggatttgg atgactgcat ctttgaacaa taaatgattt aaatcaggta     1980
tggctgccga tggttatctt ccagattggc tcgaggacac tctctctgaa ggaataagac     2040
agtggtggaa gctcaaacct ggcccaccac caccaaagcc cgcagagcgg cataaggacg     2100
```

acagcagggg tcttgtgctt cctgggtaca agtacctcgg acccttcaac ggactcgaca 2160

agggagagcc ggtcaacgag gcagacgccg cggccctcga gcacgacaaa gcctacgacc 2220

ggcagctcga cagcggagac aacccgtacc tcaagtacaa ccacgccgac gcggagtttc 2280

aggagcgcct taaagaagat acgtcttttg ggggcaacct cggacgagca gtcttccagg 2340

cgaaaaagag ggttcttgaa cctctgggcc tggttgagga acctgttaag accgctccgg 2400

gaaaaaagag gccggtagag cactctcctg tggagccaga ctcctcctcg ggaaccggaa 2460

aggcgggcca gcagcctgca agaaaaagat tgaattttgg tcagactgga gacgcagact 2520

cagtacctga cccccagcct ctcggacagc caccagcagc cccctctggt ctgggaacta 2580

atacgatggc tacaggcagt ggcgcaccaa tggcagacaa taacgagggc gccgacggag 2640

tgggtaattc ctcgggaaat tggcattgcg attccacatg gatgggcgac agagtcatca 2700

ccaccagcac ccgaacctgg gccctgccca cctacaacaa ccacctctac aaacaaattt 2760

ccagccaatc aggagcctcg aacgacaatc actactttgg ctacagcacc ccttgggggt 2820

attttgactt caacagattc cactgccact tttcaccacg tgactggcaa agactcatca 2880

acaacaactg gggattccga cccaagagac tcaacttcaa gctctttaac attcaagtca 2940

aagaggtcac gcagaatgac ggtacgacga cgattgccaa taaccttacc agcacggttc 3000

aggtgtttac tgactcggag taccagctcc cgtacgtcct cggctcggcg catcaaggat 3060

gcctcccgcc gttcccagca gacgtcttca tggtgccaca gtatggatac ctcaccctga 3120

acaacgggag tcaggcagta ggacgctctt cattttactg cctggagtac tttccttctc 3180

agatgctgcg taccggaaac aactttacct tcagctacac ttttgaggac gttcctttcc 3240

acagcagcta cgctcacagc cagagtctgg accgtctcat gaatcctctc atcgaccagt 3300

acctgtatta cttgagcaga acaaacactc caagtggaac caccacgcag tcaaggcttc 3360

agttttctca ggccggagcg agtgacattc gggaccagtc taggaactgg cttcctggac 3420

cctgttaccg ccagcagcga gtatcaaaga catctgcgga taacaacaac agtgaatact 3480

cgtggactgg agctaccaag taccacctca atggcagaga ctctctggtg aatccgggcc 3540

cggccatggc aagccacaag gacgatgaag aaaagttttt tcctcagagc ggggttctca 3600

tctttgggaa gcaaggctca gagaaaacaa atgtggacat tgaaaaggtc atgattacag 3660

acgaagagga aatcaggaca accaatcccg tggctacgga gcagtatggt tctgtatcta 3720

ccaacctcca gagaggcaac agacaagcag ctaccgcaga tgtcaacaca caaggcgttc 3780

ttccaggcat ggtctggcag gacagagatg tgtaccttca ggggcccatc tgggcaaaga 3840

ttccacacac ggacggacat tttcacccct ctcccctcat gggtggattc ggacttaaac 3900

accctcctcc acagattctc atcaagaaca ccccggtacc tgcgaatcct tcgaccacct 3960

tcagtgcggc aaagtttgct tccttcatca cacagtactc cacgggacag gtcagcgtgg 4020

agatcgagtg ggagctgcag aaggaaaaca gcaaacgctg gaatcccgaa attcagtaca 4080

cttccaacta caacaagtct gttaatgtgg actttactgt ggacactaat ggcgtgtatt 4140

cagagcctcg ccccattggc accagatacc tgactcgtaa tctgtaattg cttgttaatc 4200

aataaaccgt ttaattcgtt tcagttgaac tttggtctct gcgtatttct ttcttatcta 4260

gtttccatgg ctacgtagat aagtagcatg gcgggttaat cattaactac agcccgggcg 4320

tttaaacagc gggcggaggg gtggagtcgt gacgtgaatt acgtcatagg gttagggagg 4380

tcctgtatta gaggtcacgt gagtgttttg cgacattttg cgacaccatg tggtctcgct 4440

ggggggggag gcccgagtga gcacgcaggg tctccatttt gaagcgggag gtttgaacga 4500

gcgctggcgc gctcactggc cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt 4560

acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa tagcgaagag 4620

gcccgcaccg atcgcccttc cc 4642

<210> 6
<211> 4642
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 6
gctctagaac tagtggatcc cccggaagat cagaagttcc tattccgaag ttcctattct 60

ctagaaagta taggaacttc tgatctgcgc agccgccatg ccggggtttt acgagattgt 120

gattaaggtc cccagcgacc ttgacgagca tctgcccggc atttctgaca gctttgtgaa 180

ctgggtggcc gagaaggaat gggagttgcc gccagattct gacatggatc tgaatctgat 240

tgagcaggca cccctgaccg tggccgagaa gctgcagcgc gactttctga cggaatggcg 300

ccgtgtgagt aaggccccgg aggccctttt ctttgtgcaa tttgagaagg gagagagcta 360

cttccacatg cacgtgctcg tggaaaccac cggggtgaaa tccatggttt tgggacgttt 420

cctgagtcag attcgcgaaa aactgattca gagaatttac cgcgggatcg agccgacttt 480

gccaaactgg ttcgcggtca caaagaccag aaatggcgcc ggaggcggga acaaggtggt 540

ggatgagtgc tacatcccca attacttgct ccccaaaacc cagcctgagc tccagtgggc 600

gtggactaat atggaacagt atttaagcgc ctgtttgaat ctcacggagc gtaaacggtt 660

ggtggcgcag catctgacgc acgtgtcgca gacgcaggag cagaacaaag agaatcagaa 720

tcccaattct gatgcgccgg tgatcagatc aaaaacttca gccaggtaca tggagctggt 780

cgggtggctc gtggacaagg ggattacctc ggagaagcag tggatccagg aggaccaggc 840

ctcatacatc tccttcaatg cggcctccaa ctcgcggtcc caaatcaagg ctgccttgga 900

caatgcggga aagattatga gcctgactaa aaccgccccc gactacctgg tgggccagca 960

gcccgtggag gacatttcca gcaatcggat ttataaaatt ttggaactaa acgggtacga 1020

tccccaatat gcggcttccg tctttctggg atgggccacg aaaaagttcg gcaagaggaa 1080

caccatctgg ctgtttgggc ctgcaactac cgggaagacc aacatcgcgg aggccatagc 1140

ccacactgtg cccttctacg ggtgcgtaaa ctggaccaat gagaactttc ccttcaacga 1200

ctgtgtcgac aagatggtga tctggtggga ggaggggaag atgaccgcca aggtcgtgga 1260

gtcggccaaa gccattctcg gaggaagcaa ggtgcgcgtg gaccagaaat gcaagtcctc 1320

ggcccagata gacccgactc ccgtgatcgt cacctccaac accaacatgt gcgccgtgat 1380

tgacgggaac tcaacgacct tcgaacacca gcagccgttg caagaccgga tgttcaaatt 1440

tgaactcacc cgccgtctgg atcatgactt tgggaaggtc accaagcagg aagtcaaaga 1500

ctttttccgg tgggcaaagg atcacgtggt tgaggtggag catgaattct acgtcaaaaa 1560

gggtggagcc aagaaaagac ccgcccccag tgacgcagat ataagtgagc ccaaacgggt 1620

gcgcgagtca gttgcgcagc catcgacgtc agacgcggaa gcttcgatca actacgcaga 1680

caggtaccaa aacaaatgtt ctcgtcacgt gggcatgaat ctgatgctgt ttccctgcag 1740

acaatgcgag agaatgaatc agaattcaaa tatctgcttc actcacggac agaaagactg 1800

tttagagtgc tttcccgtgt cagaatctca acccgtttct gtcgtcaaaa aggcgtatca 1860

gaaactgtgc tacattcatc atatcatggg aaaggtgcca gacgcttgca ctgcctgcga 1920

tctggtcaat gtggatttgg atgactgcat ctttgaacaa taaatgattt aaatcaggtc 1980

tcgctgccga tggttatctt ccagattggc tcgaggacac tctctctgaa ggaataagac 2040

agtggtggaa gctcaaacct ggcccaccac caccaaagcc cgcagagcgg cataaggacg 2100

acagcagggg tcttgtgctt cctgggtaca agtacctcgg acccttcaac ggactcgaca 2160

agggagagcc ggtcaacgag gcagacgccg cggccctcga gcacgacaaa gcctacgacc 2220

ggcagctcga cagcggagac aacccgtacc tcaagtacaa ccacgccgac gcggagtttc 2280

aggagcgcct taaagaagat acgtcttttg ggggcaacct cggacgagca gtcttccagg 2340

cgaaaaagag ggttcttgaa cctctgggcc tggttgagga acctgttaag accgctccgg 2400

gaaaaaagag gccggtagag cactctcctg tggagccaga ctcctcctcg ggaaccggaa 2460

aggcgggcca gcagcctgca agaaaaagat tgaattttgg tcagactgga gacgcagact 2520

cagtacctga cccccagcct ctcggacagc caccagcagc cccctctggt ctgggaacta 2580

atacgatggc tacaggcagt ggcgcaccaa tggcagacaa taacgagggc gccgacggag 2640

tgggtaattc ctcgggaaat tggcattgcg attccacatg gatgggcgac agagtcatca 2700

ccaccagcac ccgaacctgg gccctgccca cctacaacaa ccacctctac aaacaaattt 2760

ccagccaatc aggagcctcg aacgacaatc actactttgg ctacagcacc ccttgggggt 2820

attttgactt caacagattc cactgccact tttcaccacg tgactggcaa agactcatca 2880

acaacaactg gggattccga cccaagagac tcaacttcaa gctctttaac attcaagtca 2940

aagaggtcac gcagaatgac ggtacgacga cgattgccaa taaccttacc agcacggttc 3000

aggtgtttac tgactcggag taccagctcc cgtacgtcct cggctcggcg catcaaggat 3060

gcctcccgcc gttcccagca gacgtcttca tggtgccaca gtatggatac ctcaccctga 3120

acaacgggag tcaggcagta ggacgctctt cattttactg cctggagtac tttccttctc 3180

agatgctgcg taccggaaac aactttacct tcagctacac ttttgaggac gttcctttcc 3240

acagcagcta cgctcacagc cagagtctgg accgtctcat gaatcctctc atcgaccagt 3300

acctgtatta cttgagcaga acaaacactc caagtggaac caccacgcag tcaaggcttc 3360

agttttctca ggccggagcg agtgacattc gggaccagtc taggaactgg cttcctggac 3420

cctgttaccg ccagcagcga gtatcaaaga catctgcgga taacaacaac agtgaatact 3480

cgtggactgg agctaccaag taccacctca atggcagaga ctctctggtg aatccgggcc 3540

cggccatggc aagccacaag gacgatgaag aaaagttttt tcctcagagc ggggttctca 3600

tctttgggaa gcaaggctca gagaaaacaa atgtggacat tgaaaaggtc atgattacag 3660

acgaagagga aatcaggaca accaatcccg tggctacgga gcagtatggt tctgtatcta 3720

ccaacctcca gagaggcaac agacaagcag ctaccgcaga tgtcaacaca caaggcgttc 3780

ttccaggcat ggtctggcag gacagagatg tgtaccttca ggggcccatc tgggcaaaga 3840

ttccacacac ggacggacat tttcacccct ctcccctcat gggtggattc ggacttaaac 3900

accctcctcc acagattctc atcaagaaca ccccggtacc tgcgaatcct tcgaccacct 3960

tcagtgcggc aaagtttgct tccttcatca cacagtactc cacgggacag gtcagcgtgg 4020

agatcgagtg ggagctgcag aaggaaaaca gcaaacgctg gaatcccgaa attcagtaca 4080

cttccaacta caacaagtct gttaatgtgg actttactgt ggacactaat ggcgtgtatt 4140

cagagcctcg ccccattggc accagatacc tgactcgtaa tctgtaattg cttgttaatc 4200

aataaaccgt ttaattcgtt tcagttgaac tttggtctct gcgtatttct ttcttatcta 4260

gtttccatgg ctacgtagat aagtagcatg gcgggttaat cattaactac agcccgggcg 4320

tttaaacagc gggcggaggg gtggagtcgt gacgtgaatt acgtcatagg gttagggagg 4380

tcctgtatta gaggtcacgt gagtgttttg cgacattttg cgacaccatg tggtctcgct 4440

ggggggggag gcccgagtga gcacgcaggg tctccatttt gaagcgggag gtttgaacga 4500

gcgctggcgc gctcactggc cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt 4560

acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa tagcgaagag 4620

gcccgcaccg atcgcccttc cc 4642

<210> 7
<211> 2877
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 7
gttgacattg attattgact agttattaat agtaatcaat tacggggtca ttagttcata 60

gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 120

ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180

ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240

atcaagtgta tcatatgcca agtacgcccc ctattgacgt caatgacggt aaatggcccg 300

cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360

tattagtcat cgctattacc atggtgatgc ggttttggca gtacatcaat gggcgtggat 420

agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt 480

tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc ccattgacgc 540

aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctctc tggctaacta 600

gagaacccac tgcttactgg cttatcgaaa ttaatacgac tcactatagg gagacccaag 660

ctggctagca tggctgccga tggttatctt ccagattggc tcgaggacac tctctctgaa 720

ggaataagac agtggtggaa gctcaaacct ggcccaccac caccaaagcc cgcagagcgg 780

cataaggacg acagcagggg tcttgtgctt cctgggtaca agtacctcgg acccttcaac 840

ggactcgaca agggagagcc ggtcaacgag gcagacgccg cggccctcga gcacgacaaa 900

gcctacgacc ggcagctcga cagcggagac aacccgtacc tcaagtacaa ccacgccgac 960

gcggagtttc aggagcgcct taaagaagat acgtcttttg ggggcaacct cggacgagca 1020

gtcttccagg cgaaaaagag ggttcttgaa cctctgggcc tggttgagga acctgttaag 1080

accgctccgg gaaaaaagag gccggtagag cactctcctg tggagccaga ctcctcctcg 1140

ggaaccggaa aggcgggcca gcagcctgca agaaaaagat tgaattttgg tcagactgga 1200

gacgcagact cagtacctga cccccagcct ctcggacagc caccagcagc cccctctggt 1260

ctcggaacta ataccctcgc tacaggcagt ggcgcaccac tcgcagacaa taacgagggc 1320

gccgacggag tgggtaattc ctcgggaaat tggcattgcg attccacatg gctcggcgac 1380

agagtcatca ccaccagcac ccgaacctgg gccctgccca cctacaacaa ccacctctac 1440

aaacaaattt ccagccaatc aggagcctcg aacgacaatc actactttgg ctacagcacc 1500

ccttgggggt attttgactt caacagattc cactgccact tttcaccacg tgactggcaa 1560

agactcatca acaacaactg gggattccga cccaagagac tcaacttcaa gctctttaac 1620

attcaagtca aagaggtcac gcagaatgac ggtacgacga cgattgccaa taaccttacc 1680

agcacggttc aggtgtttac tgactcggag taccagctcc cgtacgtcct cggctcggcg 1740

catcaaggat gcctcccgcc gttcccagca gacgtcttca tggtgccaca gtatggatac 1800

ctcaccctga acaacgggag tcaggcagta ggacgctctt cattttactg cctggagtac 1860

tttccttctc agatgctgcg taccggaaac aactttacct tcagctacac ttttgaggac 1920

gttcctttcc acagcagcta cgctcacagc cagagtctgg accgtctcat gaatcctctc 1980

atcgaccagt acctgtatta cttgagcaga acaaacactc caagtggaac caccacgcag 2040

tcaaggcttc agttttctca ggccggagcg agtgacattc gggaccagtc taggaactgg 2100

cttcctggac cctgttaccg ccagcagcga gtatcaaaga catctgcgga taacaacaac 2160

agtgaatact cgtggactgg agctaccaag taccacctca atggcagaga ctctctggtg 2220

aatccgggcc cggccatggc aagccacaag gacgatgaag aaaagttttt tcctcagagc 2280

ggggttctca tctttgggaa gcaaggctca gagaaaacaa atgtggacat tgaaaaggtc 2340

atgattacag acgaagagga aatcaggaca accaatcccg tggctacgga gcagtatggt 2400

tctgtatcta ccaacctcca gagaggcaac agacaagcag ctaccgcaga tgtcaacaca 2460

caaggcgttc ttccaggcat ggtctggcag gacagagatg tgtaccttca ggggcccatc 2520

tgggcaaaga ttccacacac ggacggacat tttcaccccct ctcccctcat gggtggattc 2580

ggacttaaac accctcctcc acagattctc atcaagaaca ccccggtacc tgcgaatcct 2640

tcgaccacct tcagtgcggc aaagtttgct tccttcatca cacagtactc cacgggacag 2700

gtcagcgtgg agatcgagtg ggagctgcag aaggaaaaca gcaaacgctg gaatcccgaa 2760

attcagtaca cttccaacta caacaagtct gttaatgtgg actttactgt ggacactaat 2820

ggcgtgtatt cagagcctcg ccccattggc accagatacc tgactcgtaa tctgtaa 2877

<210> 8
<211> 2592
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 8
gttgacattg attattgact agttattaat agtaatcaat tacggggtca ttagttcata 60

gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 120

ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180

ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240

atcaagtgta tcatatgcca agtacgcccc ctattgacgt caatgacggt aaatggcccg 300

cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg 360

tattagtcat cgctattacc atggtgatgc ggttttggca gtacatcaat gggcgtggat 420

agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt 480

tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc ccattgacgc 540

aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctctc tggctaacta 600

gagaacccac tgcttactgg cttatcgaaa ttaatacgac tcactatagg gagacccaag 660

ctggctagca tggctccggg aaaaaagagg ccggtagagc actctcctgt ggagccagac 720

tcctcctcgg gaaccggaaa ggcgggccag cagcctgcaa gaaaaagatt gaattttggt 780

cagactggag acgcagactc agtacctgac ccccagcctc tcggacagcc accagcagcc 840

ccctctggtc tcggaactaa taccctcgct acaggcagtg gcgcaccact cgcagacaat 900

aacgagggcg ccgacggagt gggtaattcc tcgggaaatt ggcattgcga ttccacatgg 960

ctcggcgaca gagtcatcac caccagcacc cgaacctggg ccctgcccac ctacaacaac 1020

cacctctaca aacaaatttc cagccaatca ggagcctcga acgacaatca ctactttggc 1080

tacagcaccc cttgggggta ttttgacttc aacagattcc actgccactt ttcaccacgt 1140

gactggcaaa gactcatcaa caacaactgg ggattccgac ccaagagact caacttcaag 1200

ctctttaaca ttcaagtcaa agaggtcacg cagaatgacg gtacgacgac gattgccaat 1260

aaccttacca gcacggttca ggtgtttact gactcggagt accagctccc gtacgtcctc 1320

ggctcggcgc atcaaggatg cctcccgccg ttcccagcag acgtcttcat ggtgccacag 1380

tatggatacc tcaccctgaa caacgggagt caggcagtag gacgctcttc attttactgc 1440

ctggagtact ttccttctca gatgctgcgt accggaaaca actttacctt cagctacact 1500

tttgaggacg ttcctttcca cagcagctac gctcacagcc agagtctgga ccgtctcatg 1560

aatcctctca tcgaccagta cctgtattac ttgagcagaa caaacactcc aagtggaacc 1620

accacgcagt caaggcttca gttttctcag gccggagcga gtgacattcg ggaccagtct 1680

aggaactggc ttcctggacc ctgttaccgc cagcagcgag tatcaaagac atctgcggat 1740

aacaacaaca gtgaatactc gtggactgga gctaccaagt accacctcaa tggcagagac 1800

tctctggtga atccgggccc ggccatggca agccacaagg acgatgaaga aaagtttttt 1860

cctcagagcg gggttctcat ctttgggaag caaggctcag agaaaacaaa tgtggacatt 1920

gaaaaggtca tgattacaga cgaagaggaa atcaggacaa ccaatcccgt ggctacggag 1980

cagtatggtt ctgtatctac caacctccag agaggcaaca gacaagcagc taccgcagat 2040

gtcaacacac aaggcgttct tccaggcatg gtctggcagg acagagatgt gtaccttcag 2100

gggcccatct gggcaaagat tccacacacg gacggacatt ttcacccctc tcccctcatg 2160

ggtggattcg gacttaaaca ccctcctcca cagattctca tcaagaacac cccggtacct 2220

gcgaatcctt cgaccacctt cagtgcggca aagtttgctt ccttcatcac acagtactcc 2280

acgggacagg tcagcgtgga gatcgagtgg gagctgcaga aggaaaacag caaacgctgg 2340

aatcccgaaa ttcagtacac ttccaactac aacaagtctg ttaatgtgga ctttactgtg 2400

gacactaatg gcgtgtattc agagcctcgc cccattggca ccagatacct gactcgtaat 2460

ctgtaattgc ttgttaatca ataaaccgtt taattcgttt cagttgaact ttggtctctg 2520

cgtatttctt tcttatctag tttccatggc tacgtagata agtagcatgg cgggttaatc 2580

attaactaca gc 2592

<210> 9
<211> 5595
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 9
gttgacattg attattgact agttattaat agtaatcaat tacggggtca ttagttcata 60

gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgcct ggctgaccgc 120

ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 180

ggactttcca ttgacgtcaa tgggtggact atttacggta aactgcccac ttggcagtac 240

atcaagtgta tcatatgcca agtacgcccc ctattgacgt caatgacggt aaatggcccg 300

```
cctggcatta tgcccagtac atgaccttat gggactttcc tacttggcag tacatctacg      360

tattagtcat cgctattacc atggtgatgc ggttttggca gtacatcaat gggcgtggat      420

agcggtttga ctcacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt      480

tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgcc ccattgacgc      540

aaatgggcgg taggcgtgta cggtgggagg tctatataag cagagctctc tggctaacta      600

gagaacccac tgcttactgg cttatcgaaa ttaatacgac tcactatagg gagacccaag      660

ctggctagca tggctgccga tggttatctt ccagattggc tcgaggacac tctctctgaa      720

ggaataagac agtggtggaa gctcaaacct ggcccaccac caccaaagcc cgcagagcgg      780

cataaggacg acagcagggg tcttgtgctt cctgggtaca agtacctcgg acccttcaac      840

ggactcgaca agggagagcc ggtcaacgag gcagacgccg cggccctcga gcacgacaaa      900

gcctacgacc ggcagctcga cagcggagac aacccgtacc tcaagtacaa ccacgccgac      960

gcggagtttc aggagcgcct taaagaagat acgtcttttg ggggcaacct cggacgagca     1020

gtcttccagg cgaaaaagag ggttcttgaa cctctgggcc tggttgagga acctgttaag     1080

accgctccgg gaaaaaagag gccggtagag cactctcctg tggagccaga ctcctcctcg     1140

ggaaccggaa aggcgggcca gcagcctgca agaaaaagat tgaattttgg tcagactgga     1200

gacgcagact cagtacctga cccccagcct ctcggacagc caccagcagc cccctctggt     1260

ctcggaacta ataccctcgc tacaggcagt ggcgcaccac tcgcagacaa taacgagggc     1320

gccgacggag tgggtaattc ctcgggaaat tggcattgcg attccacatg gctcggcgac     1380

agagtcatca ccaccagcac ccgaacctgg gccctgccca cctacaacaa ccacctctac     1440

aaacaaattt ccagccaatc aggagcctcg aacgacaatc actactttgg ctacagcacc     1500

ccttgggggt attttgactt caacagattc cactgccact tttcaccacg tgactggcaa     1560

agactcatca acaacaactg gggattccga cccaagagac tcaacttcaa gctctttaac     1620

attcaagtca aagaggtcac gcagaatgac ggtacgacga cgattgccaa taaccttacc     1680

agcacggttc aggtgtttac tgactcggag taccagctcc cgtacgtcct cggctcggcg     1740

catcaaggat gcctcccgcc gttcccagca gacgtcttca tggtgccaca gtatggatac     1800

ctcaccctga acaacgggag tcaggcagta ggacgctctt cattttactg cctggagtac     1860

tttccttctc agatgctgcg taccggaaac aactttacct tcagctacac ttttgaggac     1920

gttcctttcc acagcagcta cgctcacagc cagagtctgg accgtctcat gaatcctctc     1980

atcgaccagt acctgtatta cttgagcaga acaaacactc caagtggaac caccacgcag     2040

tcaaggcttc agttttctca ggccggagcg agtgacattc gggaccagtc taggaactgg     2100

cttcctggac cctgttaccg ccagcagcga gtatcaaaga catctgcgga taacaacaac     2160

agtgaatact cgtggactgg agctaccaag taccacctca atggcagaga ctctctggtg     2220

aatccgggcc cggccatggc aagccacaag gacgatgaag aaaagttttt tcctcagagc     2280

ggggttctca tctttgggaa gcaaggctca gagaaaacaa atgtggacat tgaaaaggtc     2340
```

atgattacag acgaagagga aatcaggaca accaatcccg tggctacgga gcagtatggt 2400

tctgtatcta ccaacctcca gagaggcaac agacaagcag ctaccgcaga tgtcaacaca 2460

caaggcgttc ttccaggcat ggtctggcag gacagagatg tgtaccttca ggggcccatc 2520

tgggcaaaga ttccacacac ggacggacat tttcacccct ctcccctcat gggtggattc 2580

ggacttaaac accctcctcc acagattctc atcaagaaca ccccggtacc tgcgaatcct 2640

tcgaccacct tcagtgcggc aaagtttgct tccttcatca cacagtactc cacgggacag 2700

gtcagcgtgg agatcgagtg ggagctgcag aaggaaaaca gcaaacgctg gaatcccgaa 2760

attcagtaca cttccaacta caacaagtct gttaatgtgg actttactgt ggacactaat 2820

ggcgtgtatt cagagcctcg ccccattggc accagatacc tgactcgtaa tctgtaattg 2880

cttgttaatc aataaaccgt ttaattcgtt tcagttgaac tttggtctct gcgtatttct 2940

ttcttatcta gtttccatgg ctacgtagat aagtagcatg gcgggttaat cattaactac 3000

aacgcgttga cattgattat tgactagtta ttaatagtaa tcaattacgg ggtcattagt 3060

tcatagccca tatatggagt tccgcgttac ataacttacg gtaaatggcc cgcctggctg 3120

accgcccaac gacccccgcc cattgacgtc aataatgacg tatgttccca tagtaacgcc 3180

aatagggact ttccattgac gtcaatgggt ggactattta cggtaaactg cccacttggc 3240

agtacatcaa gtgtatcata tgccaagtac gccccctatt gacgtcaatg acggtaaatg 3300

gcccgcctgg cattatgccc agtacatgac cttatgggac tttcctactt ggcagtacat 3360

ctacgtatta gtcatcgcta ttaccatggt gatgcggttt tggcagtaca tcaatgggcg 3420

tggatagcgg tttgactcac ggggatttcc aagtctccac cccattgacg tcaatgggag 3480

tttgttttgg caccaaaatc aacgggactt tccaaaatgt cgtaacaact ccgccccatt 3540

gacgcaaatg ggcggtaggc gtgtacggtg ggaggtctat ataagcagag ctctctggct 3600

aactagagaa cccactgctt actggcttat cgaaattaat acgactcact atagggagac 3660

ccaagctggc tagcatggct ccgggaaaaa agaggccggt agagcactct cctgtggagc 3720

cagactcctc ctcgggaacc ggaaaggcgg gccagcagcc tgcaagaaaa agattgaatt 3780

ttggtcagac tggagacgca gactcagtac ctgaccccca gcctctcgga cagccaccag 3840

cagccccctc tggtctcgga actaataccc tcgctacagg cagtggcgca ccactcgcag 3900

acaataacga gggcgccgac ggagtgggta attcctcggg aaattggcat tgcgattcca 3960

catggctcgg cgacagagtc atcaccacca gcacccgaac ctgggccctg cccacctaca 4020

acaaccacct ctacaaacaa atttccagcc aatcaggagc ctcgaacgac aatcactact 4080

ttggctacag caccccttgg gggtattttg acttcaacag attccactgc cacttttcac 4140

cacgtgactg gcaaagactc atcaacaaca actggggatt ccgacccaag agactcaact 4200

tcaagctctt taacattcaa gtcaaagagg tcacgcagaa tgacggtacg acgacgattg 4260

ccaataacct taccagcacg gttcaggtgt ttactgactc ggagtaccag ctcccgtacg 4320

tcctcggctc ggcgcatcaa ggatgcctcc cgccgttccc agcagacgtc ttcatggtgc 4380

cacagtatgg atacctcacc ctgaacaacg ggagtcaggc agtaggacgc tcttcatttt 4440

actgcctgga gtactttcct tctcagatgc tgcgtaccgg aaacaacttt accttcagct 4500

acacttttga ggacgttcct ttccacagca gctacgctca cagccagagt ctggaccgtc 4560

tcatgaatcc tctcatcgac cagtacctgt attacttgag cagaacaaac actccaagtg 4620

gaaccaccac gcagtcaagg cttcagtttt ctcaggccgg agcgagtgac attcgggacc 4680

agtctaggaa ctggcttcct ggaccctgtt accgccagca gcgagtatca aagacatctg 4740

cggataacaa caacagtgaa tactcgtgga ctggagctac caagtaccac ctcaatggca 4800

gagactctct ggtgaatccg ggcccggcca tggcaagcca caaggacgat gaagaaaagt 4860

tttttcctca gagcggggtt ctcatctttg ggaagcaagg ctcagagaaa acaaatgtgg 4920

acattgaaaa ggtcatgatt acagacgaag aggaaatcag gacaaccaat cccgtggcta 4980

cggagcagta tggttctgta tctaccaacc tccagagagg caacagacaa gcagctaccg 5040

cagatgtcaa cacacaaggc gttcttccag gcatggtctg gcaggacaga gatgtgtacc 5100

ttcaggggcc catctgggca aagattccac acacggacgg acattttcac ccctctcccc 5160

tcatgggtgg attcggactt aaacaccctc ctccacagat tctcatcaag aacaccccgg 5220

tacctgcgaa tccttcgacc accttcagtg cggcaaagtt tgcttccttc atcacacagt 5280

actccacggg acaggtcagc gtggagatcg agtgggagct gcagaaggaa aacagcaaac 5340

gctggaatcc cgaaattcag tacacttcca actacaacaa gtctgttaat gtggacttta 5400

ctgtggacac taatggcgtg tattcagagc ctcgccccat tggcaccaga tacctgactc 5460

gtaatctgta attgcttgtt aatcaataaa ccgtttaatt cgtttcagtt gaactttggt 5520

ctctgcgtat ttctttctta tctagtttcc atggctacgt agataagtag catggcgggt 5580

taatcattaa ctaca 5595

<210> 10
<211> 10010
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 10
cccgtgtaaa acgacggcca gtttatctag tcagcttgat tctagctgat cgtggaccgg 60

aaggtgagcc agtgagttga ttgcagtcca gttacgctgg agtctgaggc tcgtcctgaa 120

tgatatgcga ccgccggagg gttgcgtttg agacgggcga cagatccagt cgcgctgctc 180

tcgtcgatcc gctaggacta tggttgcttt gacgtatgcg gtgtgaaata ccgcacagat 240

gcgtaaggag aaaataccgc atcaggcgcc cctgcaggca gctgcgcgct cgctcgctca 300

ctgaggccgc ccgggcaaag cccgggcgtc gggcgacctt tggtcgcccg gcctcagtga 360

gcgagcgagc gcgcagagag ggagtggcca actccatcac taggggttcc tgcggccgct 420

cggtccgcac gtggttacct acaaaatcag aaggacaggg aagggagcag tggttcacgc 480

ctgtaatccc agcaatttgg gaggccaagg tgggtagatc acctgagatt aggagttgga 540

gaccagcctg gccaatatgg tgaaaccccg tctctaccaa aaaaacaaaa attagctgag 600

cctggtcatg catgcctgga atcccaacaa ctcgggaggc tgaggcagga gaatcgcttg 660

aacccaggag gcggagattg cagtgagcca agattgtgcc actgcactcc agcttggttc 720

ccaatagacc ccgcaggccc tacaggttgt cttcccaact tgcccccttgc tccataccac 780

ccccctccac cccataatat tatagaagga cacctagtca gacaaaatga tgcaacttaa 840

ttttattagg acaaggctgg tgggcactgg agtggcaact tccagggcca ggagaggcac 900

tggggagggg tcacagggat gccacccgta gatctctcga gcagcgctgc tcgaggcaag 960

cttctgcagg tcgactctag aggatcccta cttgagctcg agatctgagt acttgtacag 1020

ctcgtccatg ccgagagtga tcccggcggc ggtcacgaac tccagcagga ccatgtgatc 1080

gcgcttctcg ttggggtctt tgctcagggc ggactgggtg ctcaggtagt ggttgtcggg 1140

cagcagcacg gggccgtcgc cgatgggggt gttctgctgg tagtggtcgg cgagctgcac 1200

gctgccgtcc tcgatgttgt ggcggatctt gaagttcacc ttgatgccgt tcttctgctt 1260

gtcggccatg atatagacgt tgtggctgtt gtagttgtac tccagcttgt gccccaggat 1320

gttgccgtcc tccttgaagt cgatgccctt cagctcgatg cggttcacca gggtgtcgcc 1380

ctcgaacttc acctcggcgc gggtcttgta gttgccgtcg tccttgaaga agatggtgcg 1440

ctcctggacg tagccttcgg gcatggcgga cttgaagaag tcgtgctgct tcatgtggtc 1500

ggggtagcgg ctgaagcact gcacgccgta ggtcagggtg gtcacgaggg tgggccaggg 1560

cacgggcagc ttgccggtgg tgcagatgaa cttcagggtc agcttgccgt aggtggcatc 1620

gccctcgccc tcgccggaca cgctgaactt gtggccgttt acgtcgccgt ccagctcgac 1680

caggatgggc accaccccgg tgaacagctc ctcgcccttg ctcaccattc agaattcaat 1740

cgatgttcga atcccaattc tttgccaaag tgatgggcca gcacacagac cagcacgttg 1800

cccaggagct gtgggaggaa gataagaggt atgaacatga ttagcaaaag ggcctagctt 1860

ggactcagaa taatccagcc ttatcccaac cataaaataa aagcagaatg gtagctggat 1920

tgtagctgct attagcaata tgaaacctct tacatcagtt acaatttata tgcagaaata 1980

tttatatgca gaaatattgc tattgcctta acccagaaat tatcactgtt attctttaga 2040

atggtgcaaa gaggcatgat acattgtatc attattgccc tgaaagaaag agattaggga 2100

aagtattaga aataagataa acaaaaaagt atattaaaag aagaaagcat tttttgtggg 2160

cctatagact ctataggcgg tacttacgtc actcttggca cggggaatcc gcgttccaat 2220

gcaccgttcc cggccgggat tcgaatccgc ggaggctgga tcggtcccgg tgtcttctat 2280

ggaggtcaaa acagcgtgga tggcgtctcc aggcgatctg acggttcact aaacgagctc 2340

tgcttatata gacctcccac cgtacacgcc taccgcccat ttgcgtcaat ggggcggagt 2400

tgttacgaca ttttggaaag tcccgttgat tttggtgcaa aacaaactcc cattgacgtc 2460

aatggggtgg agacttggaa atccccgtga gtcaaaccgc tatccacgcc cattgatgta 2520

ctgccaaaac cgcatcacca tggtaatagc gatgactaat acgtagatgt actgccaagt 2580

aggaaagtcc cataaggtca tgtactgggc ataatgccag gcgggccatt taccgtcatt 2640

gacgtcaata gggggcgtac ttggcatatg atacacttga tgtactgcca agtgggcagt 2700

ttaccgtaaa tactccaccc attgacgtca atggaaagtc cctattgacg ttactatggg 2760

aacatacgtc attattgacg tcaatgggcg ggggtcgttg ggcggtcagc caggcgggcc 2820

atttaccgta agttatgtaa cgcggaactc catatatggg ctatgaacta atgaccccgt 2880

aattgattac tattaataac tagctccacg cgtgcggccg caggaacccc tagtgatgga 2940

gttggccact ccctctctgc gcgctcgctc gctcactgag gccgggcgac caaaggtcgc 3000

ccgacgcccg ggctttgccc gggcggcctc agtgagcgag cgagcgcgca gctgcctgca 3060

ggacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc 3120

gtttttccat aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag 3180

gtggcgaaac ccgagctagc aaaaaatgcc cggagcggga cttgaacccg cacagcgcga 3240

acgccgaggg atttagagtc ccttgtgtct accgattcca ccatccgggc ggtgtttcgt 3300

cctttccaca agatatataa agccaagaaa tcgaaatact ttcaagttac ggtaagcata 3360

tgatagtcca ttttaaaaca taattttaaa actgcaaact acccaagaaa ttattacttt 3420

ctacgtcacg tattttgtac taatatcttt gtgtttacag tcaaattaat tctaattatc 3480

tctctaacag ccttgtatcg tatatgcaaa tatgaaggaa tcatgggaaa taggccctct 3540

tcctgcccga ggtaccccta gcaaaaaatg cccggagcgg gacttgaacc cgcacagcgc 3600

gaacgccgag ggatttagag tcccttgtgt ctaccgattc caccatccgg gcggtgtttc 3660

gtcctttcca caagatatat aaagccaaga aatcgaaata ctttcaagtt acggtaagca 3720

tatgatagtc cattttaaaa cataatttta aaactgcaaa ctacccaaga aattattact 3780

ttctacgtca cgtattttgt actaatatct ttgtgtttac agtcaaatta attctaatta 3840

tctctctaac agccttgtat cgtatatgca aatatgaagg aatcatggga aataggccct 3900

cttcctgccc gaggtacccc tagcaaaaaa tgcccggagc gggacttgaa cccgcacagc 3960

gcgaacgccg agggatttag agtcccttgt gtctaccgat tccaccatcc gggcggtgtt 4020

tcgtcctttc cacaagatat ataaagccaa gaaatcgaaa tactttcaag ttacggtaag 4080

catatgatag tccattttaa aacataattt taaaactgca aactacccaa gaaattatta 4140

ctttctacgt cacgtatttt gtactaatat ctttgtgttt acagtcaaat taattctaat 4200

tatctctcta acagccttgt atcgtatatg caaatatgaa ggaatcatgg gaaataggcc 4260

ctcttcctgc ccgaggtacc cctagcaaaa aatgcccgga gcgggacttg aacccgcaca 4320

gcgcgaacgc cgagggattt agagtccctt gtgtctaccg attccaccat ccgggcggtg 4380

tttcgtcctt tccacaagat atataaagcc aagaaatcga aatactttca agttacggta 4440

agcatatgat agtccatttt aaaacataat tttaaaactg caaactaccc aagaaattat 4500

tactttctac gtcacgtatt ttgtactaat atctttgtgt ttacagtcaa attaattcta 4560

attatctctc taacagcctt gtatcgtata tgcaaatatg aaggaatcat gggaaatagg 4620

ccctcttcct gcccgaggta cccctaggtg cgctgcttcg cgatgtacgg gccagatata 4680

cgcgttgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 4740

atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 4800

cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 4860

tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 4920

tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 4980

ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 5040

acgtattagt catcgctatt accatggtga tgcggttttg gcagtacatc aatgggcgtg 5100

gatagcggtt tgactcacgg ggatttccaa gtctccaccc cattgacgca aatgggcggt 5160

aggcgtgtac ggtgggaggt ctatataagc agagctctct ggctaactag agaacccact 5220

gcttactggc ttatcgaaat taatacgact cactataggg agacccaagc tggctagcgt 5280

ttaaacttaa gcttgccgcc accatggaag agcaataccg cccggaagag atagaatcca 5340

aagtacagct tcattgggat gagaagcgca catttgaagt aaccgaagac gagagcaaag 5400

agaagtatta ctgcctgtct atcctcccct atccttctgg tcgactacac atgggccacg 5460

tacgtaacta caccatcggt gacgtgatcg cccgctacca gcgtatgctg ggcaaaaacg 5520

tcctgcagcc gatcggctgg gacgcgtttg gtctgcctgc ggaaggcgcg gcggtgaaaa 5580

acaacaccgc tccggcaccg tggacgtacg acaacatcgc gtatatgaaa aaccagctca 5640

aaatgctggg ctttggttat gactggagcc gcgagctggc aacctgtacg ccggaatact 5700

accgttggga acagaaattc ttcaccgagc tgtataaaaa aggcctggta tataagaaga 5760

cttctgcggt caactggtgt ccgaacgacc agaccgtact ggcgaacgaa caagttatcg 5820

acggctgctg ctggcgctgc gataccaaag ttgaacgtaa agagatcccg cagtggttta 5880

tcaaaatcac tgcttacgct gacgagctgc tcaacgatct ggataaactg gatcactggc 5940

cagacaccgt taaaaccatg cagcgtaact ggatcggtcg ttccgaaggc gtggagatca 6000

ccttcaacgt taacgactat gacaacacgc tgaccgttta cactacccgc ccggacgcgt 6060

ttatgggttg tacctacctg gcggtagctg cgggtcatcc gctggcgcag aaagcggcgg 6120

aaaataatcc tgaactggcg gcctttattg acgaatgccg taacaccaaa gttgccgaag 6180

ctgaaatggc gacgatggag aaaaaaggcg tcgatactgg ctttaaagcg gttcacccat 6240

taacgggcga agaaattccc gtttgggcag caaacttcgt attgatggag tacggcacgg 6300

gcgcagttat ggcggtaccg gggcacgacc agcgcgacta cgagtttgcc tctaaatacg 6360

gcctgaacat caaaccggtt atcctggcag ctgacggctc tgagccagat ctttctcagc 6420

aagccctgac tgaaaaaggc gtgctgttca actctggcga gttcaacggt cttgaccatg 6480

aagcggcctt caacgccatc gccgataaac tgactgcgat gggcgttggc gagcgtaaag 6540

tgaactaccg cctgcgcgac tggggtgttt cccgtcagcg ttactggggc gcgccgattc 6600

cgatggtgac tctagaagac ggtaccgtaa tgccgacccc ggacgaccag ctgccggtga 6660

tcctgccgga agatgtggta atggacggca ttaccagccc gattaaagca gatccggagt 6720

gggcgaaaac taccgttaac ggtatgccag cactgcgtga aaccgacact ttcgacacct 6780

ttatggagtc ctcctggatc tatgcgcgct acacttgccc gcagtacaaa gaaggtatgc 6840

tggattccga agcggctaac tactggctgc cggtggatat cgctattggt ggtattgaac 6900

acgccattat gggtctgctc tacttccgct tcttccacaa actgatgcgt gatgcaggca 6960

tggtgaactc tgacgaacca gcgaaacagt tgctgtgtca gggtatggtg ctggcagatg 7020

ccttctacta tgttggcgaa aacggcgaac gtaactgggt ttccccggtt gatgctatcg 7080

ttgaacgtga cgagaaaggc cgtatcgtga aagcgaaaga tgcggcaggc catgaactgg 7140

tttataccgg catgagcaaa atgtccaagt cgaagaacaa cggtatcgac ccgcaggtga 7200

tggttgaacg ttacggcgcg gacaccgttc gtctgtttat gatgtttgct tctccggctg 7260

atatgactct cgaatggcag gaatccggtg tggaaggggc taaccgcttc ctgaaacgtg 7320

tctggaaact ggtttacgag cacacagcaa aaggtgatgt tgcggcactg aacgttgatg 7380

cgctgactga aaatcagaaa gcgctgcgtc gcgatgtgca taaaacgatc gctaaagtga 7440

ccgatgatat cggccgtcgt cagaccttca acaccgcaat tgcggcgatt atggagctga 7500

tgaacaaact ggcgaaagca ccaaccgatg gcgagcagga tcgcgctctg atgcaggaag 7560

cactgctggc cgttgtccgt atgcttaacc cgttcacccc gcacatctgc ttcacgctgt 7620

ggcaggaact gaaaggcgaa ggcgatatcg acaacgcgcc gtggccggtt gctgacgaaa 7680

aagcgatggt ggaagactcc acgctggtcg tggtgcaggt taacggtaaa gtccgtgcca 7740

aaatcaccgt tccggtggac gcaacggaag aacaggttcg cgaacgtgct ggccaggaac 7800

atctggtagc aaaatatctt gatggcgtta ctgtacgtaa agtgatttac gtaccaggta 7860

aactcctcaa tctggtcgtt ggcgggcccg tttaactcga gtctagaggg cccgtttaaa 7920

cccgctgatc agcctcgact gtgccttcta gttgccagcc atctgttgtt tgcccctccc 7980

ccgtgccttc cttgaccctg gaaggtgcca ctcccactgt cctttcctaa taaaatgagg 8040

aaattgcatc gcattgtctg agtaggtgtc attctattct gggggggtggg gtggggcagg 8100

acagcaaggg ggaggattgg gaagacaata gcaggcatgc tggggatgcg gtgggctcta 8160

tggcttctga ggcggaaaga accctagggg tgcgagcgga tcgagcagtg tcgatcacta 8220

ctggaccgcg agctgtgctg cgacccgtga tcttacggca ttatacgtat gatcggtcca 8280

cgatcagcta gattatctag tcagcttgat gtcatagctg tttcctgagg ctcaatactg 8340

accatttaaa tcatacctga cctccatagc agaaagtcaa aagcctccga ccggaggctt 8400

ttgacttgat cggcacgtaa gaggttccaa ctttcaccat aatgaaataa gatcactacc 8460

gggcgtattt tttgagttat cgagattttc aggagctaag gaagctaaaa tgagccatat 8520

tcaacgggaa acgtcttgct tgaagccgcg attaaattcc aacatggatg ctgatttata 8580

tgggtataaa tgggctcgcg ataatgtcgg gcaatcaggt gcgacaatct atcgattgta 8640

tgggaagccc gatgcgccag agttgtttct gaaacatggc aaaggtagcg ttgccaatga 8700

tgttacagat gagatggtca ggctaaactg gctgacggaa tttatgcctc ttccgaccat 8760

caagcatttt atccgtactc ctgatgatgc atggttactc accactgcga tcccagggaa 8820

aacagcattc caggtattag aagaatatcc tgattcaggt gaaaatattg ttgatgcgct 8880

ggcagtgttc ctgcgccggt tgcattcgat tcctgtttgt aattgtcctt ttaacggcga 8940

tcgcgtattt cgtctcgctc aggcgcaatc acgaatgaat aacggtttgg ttggtgcgag 9000

tgattttgat gacgagcgta atggctggcc tgttgaacaa gtctggaaag aaatgcataa 9060

actcttgcca ttctcaccgg attcagtcgt cactcatggt gatttctcac ttgataacct 9120

tatttttgac gaggggaaat taataggttg tattgatgtt ggacgagtcg gaatcgcaga 9180

ccgataccag gatcttgcca tcctatggaa ctgcctcggt gagttttctc cttcattaca 9240

gaaacggctt tttcaaaaat atggtattga taatcctgat atgaataaat tgcagtttca 9300

cttgatgctc gatgagtttt tctaatgagg acctaaatgt aatcacctgg ctcaccttcg 9360

ggtgggcctt tctgcgttgc tggcgttttt ccataggctc cgccccctg acgagcatca 9420

caaaaatcga tgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc 9480

gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata 9540

cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 9600

tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca 9660

gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg taagacacga 9720

cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg 9780

tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa cagtatttgg 9840

tatctgcgct ctgctgaagc cagttacctc ggaaaaagag ttggtagctc ttgatccggc 9900

aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat tacgcgcaga 9960

aaaaaaggat ctcaagaaga tcctttgatt ttctaccgaa gaaaggccca 10010

<210> 11
<211> 10183
<212> DNA
<213> Artificial Sequence

<220>
<221> source
<223> /note="Description of Artificial Sequence: Synthetic
polynucleotide"

<400> 11
cccgtgtaaa acgacggcca gtttatctag tcagcttgat tctagctgat cgtggaccgg 60

aaggtgagcc agtgagttga ttgcagtcca gttacgctgg agtctgaggc tcgtcctgaa 120

tgatatgcga ccgccggagg gttgcgtttg agacgggcga cagatccagt cgcgctgctc 180

tcgtcgatcc gctagggcgg ccgcaaatac ctgcaggcag ctgcgcgctc gctcgctcac 240

tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt ggtcgcccgg cctcagtgag 300

cgagcgagcg cgcagagagg gagtggccaa ctccatcact aggggttcct gcggccgctc 360

ggtccgcacg tggttaccta caaaatcaga aggacaggga agggagcagt ggttcacgcc 420

tgtaatccca gcaatttggg aggccaaggt gggtagatca cctgagatta ggagttggag 480

```
accagcctgg ccaatatggt gaaaccccgt ctctaccaaa aaaacaaaaa ttagctgagc      540

ctggtcatgc atgcctggaa tcccaacaac tcgggaggct gaggcaggag aatcgcttga      600

acccaggagg cggagattgc agtgagccaa gattgtgcca ctgcactcca gcttggttcc      660

caatagaccc cgcaggccct acaggttgtc ttcccaactt gccccttgct ccataccacc      720

cccctccacc ccataatatt atagaaggac acctagtcag acaaaatgat gcaacttaat      780

tttattagga caaggctggt gggcactgga gtggcaactt ccagggccag gagaggcact      840

ggggaggggt cacagggatg ccacccgtag atctctcgag cagcgctgct cgaggcaagc      900

ttctgcaggt cgactctaga ggatccctac ttgagctcga gatctgagta cttgtacagc      960

tcgtccatgc cgagagtgat cccggcggcg gtcacgaact ccagcaggac catgtgatcg     1020

cgcttctcgt tggggtcttt gctcagggcg gactgggtgc tcaggtagtg gttgtcgggc     1080

agcagcacgg ggccgtcgcc gatgggggtg ttctgctggt agtggtcggc gagctgcacg     1140

ctgccgtcct cgatgttgtg gcggatcttg aagttcacct tgatgccgtt cttctgcttg     1200

tcggccatga tatagacgtt gtggctgttg tagttgtact ccagcttgtg ccccaggatg     1260

ttgccgtcct ccttgaagtc gatgcccttc agctcgatgc ggttcaccag ggtgtcgccc     1320

tcgaacttca cctcggcgcg ggtcttgtag ttgccgtcgt ccttgaagaa gatggtgcgc     1380

tcctggacgt agccttcggg catggcggac ttgaagaagt cgtgctgctt catgtggtcg     1440

gggtagcggc tgaagcactg cacgccgtag gtcagggtgg tcacgagggt gggccagggc     1500

acgggcagct tgccggtggt gcagatgaac ttcagggtca gcttgccgta ggtggcatcg     1560

ccctcgccct cgccggacac gctgaacttg tggccgttta cgtcgccgtc cagctcgacc     1620

aggatgggca ccaccccggt gaacagctcc tcgcccttgc tcaccattca gaattcaatc     1680

gatgttcgaa tcccaattct ttgccaaagt gatgggccag cacacagacc agcacgttgc     1740

ccaggagctg tgggaggaag ataagaggta tgaacatgat tagcaaaagg gcctagcttg     1800

gactcagaat aatccagcct tatcccaacc ataaaataaa agcagaatgg tagctggatt     1860

gtagctgcta ttagcaatat gaaacctctt acatcagtta caatttatat gcagaaatat     1920

ttatatgcag aaatattgct attgccttaa cccagaaatt atcactgtta ttctttagaa     1980

tggtgcaaag aggcatgata cattgtatca ttattgccct gaaagaaaga gattagggaa     2040

agtattagaa ataagataaa caaaaaagta tattaaaaga agaaagcatt ttttgtgggc     2100

ctatagactc tataggcggt acttacgtca ctcttggcac ggggaatccg cgttccaatg     2160

caccgttccc ggccgggatt cgaatccgcg gaggctggat cggtcccggt gtcttctatg     2220

gaggtcaaaa cagcgtggat ggcgtctcca ggcgatctga cggttcacta aacgagctct     2280

gcttatatag acctcccacc gtacacgcct accgcccatt tgcgtcaatg gggcggagtt     2340

gttacgacat tttggaaagt cccgttgatt ttggtgcaaa acaaactccc attgacgtca     2400

atggggtgga gacttggaaa tccccgtgag tcaaaccgct atccacgccc attgatgtac     2460

tgccaaaacc gcatcaccat ggtaatagcg atgactaata cgtagatgta ctgccaagta     2520
```

ggaaagtccc ataaggtcat gtactgggca taatgccagg cgggccattt accgtcattg 2580

acgtcaatag ggggcgtact tggcatatga tacacttgat gtactgccaa gtgggcagtt 2640

taccgtaaat actccaccca ttgacgtcaa tggaaagtcc ctattgacgt tactatggga 2700

acatacgtca ttattgacgt caatgggcgg gggtcgttgg gcggtcagcc aggcgggcca 2760

tttaccgtaa gttatgtaac gcggaactcc atatatgggc tatgaactaa tgaccccgta 2820

attgattact attaataact agctccacgc gtgcggccgc aggaacccct agtgatggag 2880

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc 2940

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgcag ctgcctgcag 3000

gctcgatccg ctcgcaccct aggtcgggca ggaagagggc ctatttccca tgattccttc 3060

atatttgcat atacgataca aggctgttag agagataatt agaattaatt tgactgtaaa 3120

cacaaagata ttagtacaaa atacgtgacg tagaaagtaa taatttcttg ggtagtttgc 3180

agttttaaaa ttatgtttta aaatggacta tcatatgctt accgtaactt gaaagtattt 3240

cgatttcttg gctttatata tcttgtggaa aggacgaaac accaggggcg tagttcaatt 3300

ggtagagcac cggtcttcaa aaccgggtgt tgggagttcg agtctctccg cccctgccat 3360

tttttgctag gtcgggcagg aagagggcct atttcccatg attccttcat atttgcatat 3420

acgatacaag gctgttagag agataattag aattaatttg actgtaaaca caaagatatt 3480

agtacaaaat acgtgacgta gaaagtaata atttcttggg tagtttgcag ttttaaaatt 3540

atgttttaaa atggactatc atatgcttac cgtaacttga aagtatttcg atttcttggc 3600

tttatatatc ttgtggaaag gacgaaacac caggggcgta gttcaattgg tagagcaccg 3660

gtcttcaaaa ccgggtgttg ggagttcgag tctctccgcc cctgccattt tttgctaggt 3720

cgggcaggaa gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc 3780

tgttagagag ataattagaa ttaatttgac tgtaaacaca aagatattag tacaaaatac 3840

gtgacgtaga aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat 3900

ggactatcat atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt 3960

gtggaaagga cgaaacacca ggggcgtagt tcaattggta gagcaccggt cttcaaaacc 4020

gggtgttggg agttcgagtc tctccgcccc tgccattttt tgctaggtcg ggcaggaaga 4080

gggcctattt cccatgattc cttcatattt gcatatacga tacaaggctg ttagagagat 4140

aattagaatt aatttgactg taaacacaaa gatattagta caaaatacgt gacgtagaaa 4200

gtaataattt cttgggtagt ttgcagtttt aaaattatgt tttaaaatgg actatcatat 4260

gcttaccgta acttgaaagt atttcgattt cttggcttta tatatcttgt ggaaaggacg 4320

aaacaccagg ggcgtagttc aattggtaga gcaccggtct tcaaaaccgg gtgttgggag 4380

ttcgagtctc tccgcccctg ccatttttg ctaggtcggg caggaagagg gcctatttcc 4440

catgattcct tcatatttgc atatacgata caaggctgtt agagagataa ttagaattaa 4500

tttgactgta aacacaaaga tattagtaca aaatacgtga cgtagaaagt aataatttct 4560

tgggtagttt gcagttttaa aattatgttt taaaatggac tatcatatgc ttaccgtaac 4620

ttgaaagtat ttcgatttct tggctttata tatcttgtgg aaaggacgaa acaccagggg 4680

cgtagttcaa ttggtagagc accggtcttc aaaaccgggt gttgggagtt cgagtctctc 4740

cgcccctgcc atttttttgct aggtcgggca ggaagagggc ctatttccca tgattccttc 4800

atatttgcat atacgataca aggctgttag agagataatt agaattaatt tgactgtaaa 4860

cacaaagata ttagtacaaa atacgtgacg tagaaagtaa taatttcttg ggtagtttgc 4920

agttttaaaa ttatgtttta aaatggacta tcatatgctt accgtaactt gaaagtattt 4980

cgatttcttg gctttatata tcttgtggaa aggacgaaac accaggggcg tagttcaatt 5040

ggtagagcac cggtcttcaa aaccgggtgt tgggagttcg agtctctccg cccctgccat 5100

tttttgctag gtcgggcagg aagagggcct atttcccatg attccttcat atttgcatat 5160

acgatacaag gctgttagag agataattag aattaatttg actgtaaaca caaagatatt 5220

agtacaaaat acgtgacgta gaaagtaata atttcttggg tagtttgcag ttttaaaatt 5280

atgttttaaa atggactatc atatgcttac cgtaacttga aagtatttcg atttcttggc 5340

tttatatatc ttgtggaaag gacgaaacac caggggcgta gttcaattgg tagagcaccg 5400

gtcttcaaaa ccgggtgttg ggagttcgag tctctccgcc cctgccattt tttgctaggt 5460

cgggcaggaa gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc 5520

tgttagagag ataattagaa ttaatttgac tgtaaacaca aagatattag tacaaaatac 5580

gtgacgtaga aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat 5640

ggactatcat atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt 5700

gtggaaagga cgaaacacca ggggcgtagt tcaattggta gagcaccggt cttcaaaacc 5760

gggtgttggg agttcgagtc tctccgcccc tgccattttt tgctaggtta ttaatagtaa 5820

tcaattacgg ggtcattagt tcatagccca tatatggagt tccgcgttac ataacttacg 5880

gtaaatggcc cgcctggctg accgcccaac gacccccgcc cattgacgtc aataatgacg 5940

tatgttccca tagtaacgcc aatagggact ttccattgac gtcaatgggt ggactattta 6000

cggtaaactg cccacttggc agtacatcaa gtgtatcata tgccaagtac gccccctatt 6060

gacgtcaatg acggtaaatg gcccgcctgg cattatgccc agtacatgac cttatgggac 6120

tttcctactt ggcagtacat ctacgtatta gtcatcgcta ttaccatggt gatgcggttt 6180

tggcagtaca tcaatgggcg tggatagcgg tttgactcac ggggatttcc aagtctccac 6240

cccattgacg tcaatgggag tttgttttgg caccaaaatc aacgggactt tccaaaatgt 6300

cgtaacaact ccgccccatt gacgcaaatg ggcggtaggc gtgtacggtg ggaggtctat 6360

ataagcagag ctctctggct aactagagaa cccactgctt actggcttat cgaaattaat 6420

acgactcact atagggagac ccaagctggc tagcgccacc atgactaagc ccatcgtttt 6480

tgctggcgca cagccctcag gtgaattgac cattggtaac tacatgggtg cgctgcgtca 6540

gtgggtaaac atgcaggatg actaccattg catttactgt atcgttgacc aacacgcgat 6600

caccgtgcgc caggatgcac agaagctgcg taaagcgacg ctggatacgc tggccttgta 6660

tctggcttgt ggtatcgatc ctgagaaaag caccattttt gttcagtccc acgtgccgga 6720

acatgcacag ttaggctggg cactgaactg ctatacctac ttcggcgaac tgagtcgcat 6780

gacgcagttt aaagataaat ctgcgcgtta tgccgagaac atcaacgctg gtctgtttga 6840

ctatccggtg ctgatggcag cggacatcct gctgtatcaa actaatctgg gtccttgtgg 6900

tgaagaccag aaacagcacc tcgaactgag ccgcgatatt gcccagcgtt tcaacgcgct 6960

gtatggcgag atctttaagg tgccggagcc gtttattccg aaatctggcg cgcgcgtaat 7020

gtcgctgctg gagccgacca agaagatgtc caagtctgac gataatcgca ataacgttat 7080

cggcctgctg gaagatccga aatcggtagt gaagaaaatc aaacgtgcgg tcactgactc 7140

cgacgagccg ccggtagttc gctacgatgt gcagaacaaa gcgggcgttt ccaacctgtt 7200

ggatatcctt tcagcggtaa cgggccagag catcccagaa ctggaaaaac agttcgaagg 7260

caagatgtat ggtcatctga aaggtgaagt ggctgatgcc gtttccggta tgctgactga 7320

attgcaggaa cgctatcacc gtttccgcaa cgatgaagcc ttcctgcaac aggtgatgaa 7380

agatggcgcg gaaaaagcca gcgcgcacgc ttcccgtacg ctaaaagcgg tgtacgaagc 7440

gattggtttt gtggcgaagc cgtaagaatt caacgcgtta agtcgacaat caacctctgg 7500

attacaaaat ttgtgaaaga ttgactggta ttcttaacta tgttgctcct tttacgctat 7560

gtggatacgc tgctttaatg cctttgtatc atgctattgc ttcccgtatg gctttcattt 7620

tctcctcctt gtataaatcc tggttgctgt ctctttatga ggagttgtgg cccgttgtca 7680

ggcaacgtgg cgtggtgtgc actgtgtttg ctgacgcaac ccccactggt tggggcattg 7740

ccaccacctg tcagctcctt tccgggactt tcgctttccc cctccctatt gccacggcgg 7800

aactcatcgc cgcctgcctt gcccgctgct ggacaggggc tcggctgttg ggcactgaca 7860

attccgtggt gttgtcgggg aaatcatcgt cctttccttg gctgctcgcc tgtgttgcca 7920

cctggattct gcgcgggacg tccttctgct acgtcccttc ggccctcaat ccagcggacc 7980

ttccttcccg cggcctgctg ccggctctgc ggcctcttcc gcgtcttcgc cttcgccctc 8040

agacgagtcg gatctccctt tgggccgcct ccccgcgtcg actttaactc gagtctagag 8100

ggcccgttta aacccgctga tcagcctcga ctgtgccttc tagttgccag ccatctgttg 8160

tttgcccctc ccccgtgcct tccttgaccc tggaaggtgc cactcccact gtcctttcct 8220

aataaaatga ggaaattgca tcgcattgtc tgagtaggtg tcattctatt ctgggggggtg 8280

gggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat gctggggatg 8340

cggtgggctc tatggactag gggtgcgagc ggatcgagca gtgtcgatca ctactggacc 8400

gcgagctgtg ctgcgacccg tgatcttacg gcattatacg tatgatcggt ccacgatcag 8460

ctagattatc tagtcagctt gatgtcatag ctgtttcctg aggctcaata ctgaccattt 8520

aaatcatacc tgacctccat agcagaaagt caaaagcctc cgaccggagg cttttgactt 8580

gatcggcacg taagaggttc caactttcac cataatgaaa taagatcact accgggcgta 8640

ttttttgagt tatcgagatt ttcaggagct aaggaagcta aaatgagcca tattcaacgg 8700

gaaacgtctt gcttgaagcc gcgattaaat tccaacatgg atgctgattt atatgggtat 8760

aaatgggctc gcgataatgt cgggcaatca ggtgcgacaa tctatcgatt gtatgggaag 8820

cccgatgcgc cagagttgtt tctgaaacat ggcaaaggta gcgttgccaa tgatgttaca 8880

gatgagatgg tcaggctaaa ctggctgacg gaatttatgc ctcttccgac catcaagcat 8940

tttatccgta ctcctgatga tgcatggtta ctcaccactg cgatcccagg gaaaacagca 9000

ttccaggtat tagaagaata tcctgattca ggtgaaaata ttgttgatgc gctggcagtg 9060

ttcctgcgcc ggttgcattc gattcctgtt tgtaattgtc cttttaacgg cgatcgcgta 9120

tttcgtctcg ctcaggcgca atcacgaatg aataacggtt tggttggtgc gagtgatttt 9180

gatgacgagc gtaatggctg gcctgttgaa caagtctgga aagaaatgca taaactcttg 9240

ccattctcac cggattcagt cgtcactcat ggtgatttct cacttgataa ccttatttt 9300

gacgagggga aattaatagg ttgtattgat gttggacgag tcggaatcgc agaccgatac 9360

caggatcttg ccatcctatg gaactgcctc ggtgagtttt ctccttcatt acagaaacgg 9420

ctttttcaaa aatatggtat tgataatcct gatatgaata aattgcagtt tcacttgatg 9480

ctcgatgagt ttttctaatg aggacctaaa tgtaatcacc tggctcacct tcgggtgggc 9540

ctttctgcgt tgctggcgtt tttccatagg ctccgccccc ctgacgagca tcacaaaaat 9600

cgatgctcaa gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcgtttccc 9660

cctggaagct ccctcgtgcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc 9720

gcctttctcc cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt 9780

tcggtgtagg tcgttcgctc caagctgggc tgtgtgcacg aacccccgt tcagcccgac 9840

cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg 9900

ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca 9960

gagttcttga agtggtggcc taactacggc tacactagaa gaacagtatt tggtatctgc 10020

gctctgctga agccagttac ctcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa 10080

ccaccgctgg tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag 10140

gatctcaaga agatcctttg attttctacc gaagaaaggc cca 10183