NOVEL LIVER TARGETING ADENO-ASSOCIATED VIRAL VECTORS

[0001] This application claims priority to U.S. Provisional Application No. 62/671,265, filed May 14, 2018, which is incorporated by reference herein in its entirety.

INCORPORATION BY REFERENCE OF MATERIAL SUBMITTED ELECTRONICALLY

[0002] This application contains, as a separate part of the disclosure, a Sequence Listing in computer-readable form which is incorporated by reference in its entirety and identified as follows: Filename: 53120_Seqlisting.txt; Size: 68,361 bytes; Created: May 10, 2019.

FIELD OF INVENTION

[0003] The invention relates to novel adeno-associated virus (AAV) capsid proteins, AAV particles comprising a novel capsid protein, polynucleotides encoding these capsid proteins and AAV vectors expressing these capsid proteins. The invention also relates to methods of making the herein described AAV vectors expressing the novel capsid proteins of the invention and associated therapeutic uses thereof.

BACKGROUND

[0004] AAV is a replication-deficient parvovirus, the single-stranded DNA genome of which is about 4.7 kb in length including 145 nucleotide inverted terminal repeat (ITRs). The nucleotide sequence of the AAV serotype 2 (AAV2) genome is presented in Srivastava *et al., J. Virol.*, 45: 555-564 (1983) as corrected by Ruffing *et al., J. Gen. Virol.*, 75: 3385-3392 (1994). Cis-acting sequences directing viral DNA replication (rep), encapsidation/packaging and host cell chromosome integration are contained within the ITRs. Three AAV promoters, p5, p19, and p40 (named for their relative map locations), drive the expression of the two AAV internal open reading frames encoding rep and cap genes. The two rep promoters (p5 and p19), coupled with the differential splicing of the single AAV intron (at nucleotides 2107 and 2227), result in the production of four rep proteins (rep 78, rep 68, rep 52, and rep 40) from the rep gene. Rep proteins possess multiple enzymatic properties which are ultimately responsible for replicating the viral genome. The cap gene is expressed from the p40 promoter and it encodes the three capsid proteins VP1, VP2, and VP3. Alternative splicing and a non-consensus translational start site are responsible for the production of the three related capsid proteins. A single consensus

polyadenylation site is located at map position 95 of the AAV genome. The life cycle and genetics of AAV are reviewed in Muzyczka, *Current Topics in Microbiology and Immunology*, 158: 97-129 (1992).

[0005] When AAV infects a human cell, the viral genome can integrate into chromosome 19 resulting in latent infection of the cell. Production of infectious virus does not occur unless the cell is infected with a helper virus (for example, adenovirus or herpesvirus). In the case of adenovirus, genes E1A, E1B, E2A, E4 and VA provide helper functions. Upon infection with a helper virus, the AAV provirus is rescued and amplified, and both AAV and adenovirus are produced.

[0006] AAV possesses unique features that make it attractive as a vaccine vector for expressing immunogenic peptides/polypeptides and as a vector for delivering foreign DNA to cells, for example, in gene therapy. AAV infection of cells in culture is non-cytopathic, and natural infection of humans and other animals is silent and asymptomatic. Moreover, AAV infects many mammalian cells allowing the possibility of targeting many different tissues *in vivo*. The AAV proviral genome is infectious as cloned DNA in plasmids which makes construction of recombinant genomes feasible. Furthermore, because the signals directing AAV replication, genome encapsidation and integration are contained within the ITRs of the AAV genome, some or all of the internal approximately 4.3 kb of the genome (encoding replication and structural capsid proteins, rep-cap) may be replaced with foreign DNA such as a gene cassette containing a promoter, a DNA of interest and a polyadenylation signal. The rep and cap proteins may be provided in trans. Another significant feature of AAV is that it is an extremely stable and hearty virus. It easily withstands the conditions used to inactivate adenovirus (56° to 65°C for several hours), making cold preservation of rAAV-vectors less critical. AAV may even be lyophilized. Finally, AAV-infected cells are not resistant to superinfection.

[0007] AAV vectors find use in numerous mammalian gene therapy applications and there is a need for new and/or modified AAV vectors and associated virus that find use in gene therapy applications. The present invention provides for novel AAV vectors expressing the novel AAV capsid proteins of the present invention, and novel, non-naturally occurring AAV virions comprising those vectors or capsid proteins.

SUMMARY OF INVENTION

[0008] The invention provides for novel AAV capsid proteins, which may be novel VP1, VP2 or VP3 capsid proteins, non-naturally occurring AAV virus comprising any of these capsid proteins, and use of such AAV virus for gene therapy applications and for use in the preparation of medicaments for gene therapy applications. In some embodiments, the AAV capsid proteins were isolated and identified from various mammalian tissues. The amino acid sequences of certain novel mammalian-derived AAV capsid VP1 proteins are set out as SEQ ID NOS:1-7, and the associated locations of the respective VP2 and VP3 sequences are also herein described. Collectively, the novel capsid proteins are referred to herein as "AAV capsid proteins of the invention."

[0009] In one embodiment, the invention provides an adeno-associated virus (AAV) having a capsid protein having an amino acid sequence that is at least 95% identical to (i) any one of SEQ ID NOS:1-7, (ii) the VP2 region of any one of SEQ ID NOS:1-7, or (iii) the VP3 region of any one of SEQ ID NOS:1-7, and further having a transgene where the transgene is composed of a heterologous gene operably linked to regulatory sequences that control expression of the heterologous gene in a host cell. In another embodiment the capsid protein has the amino acid sequence of (i) any one of SEQ ID NOS:1-7, (ii) the VP2 region of any one of SEQ ID NOS:1-7, or (iii) the VP3 region of any one of SEQ ID NOS:1-7. In yet another embodiment, the AAV has an AAV inverted terminal repeat sequence. In further embodiments, the AAV are mixed with a physiologically compatible carrier.

[0010] In another embodiment, the invention provides a method of delivering a transgene to a cell involving the step of contacting the cell with any AAV disclosed herein. In another embodiment, the invention provides a method of treating a subject from a disorder or disease associated with abnormal activity of an endogenous protein involving the step of administering to the subject an effective amount of an AAV disclosed herein where the AAV has a transgene that encodes a biologically active copy of the protein. In yet another embodiment, the methods involve delivering a transgene to a liver cell.

[0011] In further embodiment, the invention provides for a composition comprising any of the AAV of the invention for delivering a transgene to a cell, such as a liver cell. In addition, the

invention provides for use of any of the AAV of the invention for the preparation of a medicament for delivering a transgene to a cell.

[0012] In one embodiment, the invention provides for an isolated adeno-associated virus (AAV) capsid protein, wherein the capsid protein comprises (i) an amino acid sequence that is at least 95%, 96%, 97%, 98%, or 99% identical to the VP1 amino acid sequence of any one of SEQ ID NOS:1-7 or the VP2 or VP3 region of any one of SEQ ID NOS:1-7 or (ii) a VP1 amino acid sequence comprising any one of SEQ ID NOS:1-7 or the VP2 or VP3 region of any one of SEQ ID NOS:1-7. In certain embodiments, the capsid protein is linked to a heterologous amino acid sequence. The invention also provides for non-naturally occurring AAV particles having or comprising any of these capsid proteins. In certain embodiments, the non-naturally occurring AAV particle comprising any of the above described VP1, VP2 or VP3 capsid proteins comprises a nucleic acid having AAV inverted terminal repeats and a transgene comprising a heterologous gene operably linked to regulatory sequences which direct expression of the heterologous gene in a host cell. In other embodiments, the non-naturally occurring AAV particle comprising any of the VP1, VP2 or VP3 capsid sequences described herein comprises a heterologous transgene operably linked to regulatory sequences that control transgene expression in a host cell. As used herein, the terms "heterologous gene" or "heterologous regulatory sequence" means that the referenced gene or regulatory sequence is not naturally present in the AAV vector or particle and is artificially introduced therein. The term "transgene" refers to a nucleic acid that comprises both a heterologous gene and regulatory sequences that are operably linked to the heterologous gene that control expression of that gene in a host cell.

[0013] The invention also provides for a polynucleotide comprising a nucleotide sequence encoding an adeno-associated virus (AAV) capsid protein, wherein the capsid protein comprises (i) an amino acid sequence that is at least 95%, 96%, 97%, 98% or 99% identical to the VP1 amino acid sequence of any one of SEQ ID NOS:1-7 or the VP2 or VP3 region of any one of SEQ ID NOS:1-7 or (ii) a VP1 amino acid sequence comprising any one of SEQ ID NOS:1-7 or the VP2 or VP3 region of any one of SEQ ID NOS:1-7, wherein the polynucleotide is operatively linked to a heterologous regulatory control sequence. As such, it is understood that the polynucleotides of the present invention are non-naturally occurring. The invention also provides for AAV vectors comprising any of these polynucleotide sequences operably linked to a

heterologous regulatory sequence and compositions comprising these AAV vectors, including pharmaceutical compositions.

[0014] In another embodiment, the invention provides for an isolated adeno-associated virus (AAV) vector comprising a polynucleotide sequence encoding a capsid protein and a heterologous transgene sequence, wherein the capsid protein comprises (i) an amino acid sequence that is at least 95%, 96%, 97%, 98% or 99% identical to the VP1 amino acid sequence of any one of SEQ ID NOS:1-7 or the VP2 or VP3 region of any one of SEQ ID NOS:1-7 or (ii) a VP1 amino acid sequence comprising any one of SEQ ID NOS:1-7 or the VP2 or VP3 region of any one of SEQ ID NOS:1-7. The invention also provides for compositions comprising these AAV vectors, including pharmaceutical compositions.

[0015] In another embodiment, the invention provides for an adeno-associated virus (AAV) comprising a capsid protein, wherein the capsid protein comprises a functional fragment of any one of SEQ ID NOS: 1-7, and further comprising a transgene comprising a heterologous gene operably linked to a regulatory sequence that controls expression of the heterologous gene in a host cell. For example, the functional fragment comprises one or more of the variable regions (VR), the constant regions which are located between the variable regions, the GBS domain, and the GH loop of the amino acid sequence of any one of SEQ ID NO: 1-7. The invention also provides for compositions comprising these AAV vectors, including pharmaceutical compositions.

[0016] In a further embodiment, the invention provides for an adeno-associated virus comprising a capsid protein, wherein the capsid protein comprises an amino acid sequence encoded by a nucleotide sequence that hybridizes to a nucleotide sequence encoding (i) an amino acid sequence of any one of SEQ ID NO: 1-7, (ii) the VP2 region of the amino acid sequence of any one of SEQ ID NO: 1-7, or the VP3 region of the amino acid sequence of any one of SEQ ID NO: 1-7, and further comprising a transgene comprising a heterologous gene operably linked to a regulatory sequence that controls expression of the heterologous gene in a host cell. For example, the nucleotide sequence hybridized to a nucleotide sequence encoding a capsid protein or a functional fragment of a capsid protein of the invention under stringent conditions. The

invention also provides for compositions comprising these AAV vectors, including pharmaceutical compositions.

[0017] The "variable regions" refer to the nine variable regions within the VP1 sequence of an AAV capsid protein. The variable regions (VR) are referred to herein as VR I, VR II, VR III VR IV, VR V, VR VI, VR VII, VR VIII and VR IX and their respective locations in various VP1 sequences are herein described. The VR exhibit the highest sequence and structural variation within the AAV VP1 capsid sequence and may also have roles in receptor attachment, transcriptional activation of transgenes, tissue transduction and antigenicity.

[0018] The "glycan binding sequence (GBS)" or 'GBS domain" or "GBS region" refer to the amino acid sequence located between VR IV and VR V that governs the glycan binding specificity of the viral capsid. The locations of the GBS regions in various AAV VP1 amino acid sequences are herein described, and those from other AAV VP1 amino acid sequences are known in the art and/or may be routine identified.

[0019] The "GH loop" refers to a loop sequence that is flanked by β-strand G and β-strand H within the internal β-barrel of the capsid protein. The "GH loop" sequence comprises variable region VR IV through VR VIII, including the encompassed GBS sequence and all interspersed conserved backbone sequence from the donor. The locations of the GH loop regions in various AAV VP1 amino acid sequences are herein described and those from other AAV VP1 amino acid sequences may be routinely identified.

[0020] In regard to the herein described locations of the VR, GBS and GH loop regions, it is noted that the location of the N-terminal and/or C-terminal ends of those regions may vary by from up to 1 amino acid, 2 amino acids, 3 amino acids, 4 amino acids or 5 amino acids from the amino acid locations of those regions as they are explicitly described herein (particularly in Table 2). Novel capsid sequence comprising substituted VR, GBS and/or GH loop region(s) that vary from up to 5 amino acids on the N-terminal and/or C-terminal end as herein defined are encompassed by the present invention.

[0021] The invention provides for methods of producing a recombinant adeno-associated virus (AAV) particle comprising the steps of: culturing a cell that has been transfected with any of the AAV vectors of the invention and recovering recombinant AAV particle from the supernatant of

the transfected cell. In addition, the invention provides for viral particles comprising any of the viral vectors or capsid proteins of the invention and cells comprising these viral vectors.

[0022] One embodiment of the invention provides a method of producing any of the recombinant AAV described herein by culturing a viral production cell into which has been introduced a first nucleic acid vector having 5' and 3' AAV inverted terminal repeat sequences flanking a transgene having a heterologous gene operably linked to regulatory sequences that control expression of the heterologous gene in a host cell, and a second nucleic acid vector having AAV rep and cap nucleic acids sequences, wherein said cap nucleic acid sequence encodes an AAV capsid that is at least 95% identical to any of SEQ ID NOS:1-7; and recovering the AAV from the supernatant of the viral production cell culture. In another embodiment the viral production cell is a mammalian. In a preferred embodiment the insect cell is a HEK293 cell.

[0023] In another embodiment, the invention provides for methods of treating a patient suffering from a disorder or disease comprising administering to the patient an effective amount of any of the AAV vectors or virus of the invention.

[0024] In a further embodiment, the invention provides for use of any of the AAV vectors or virus of the invention for preparation of a medicament for the treatment of a disorder or disease. The invention also provides for compositions comprising any of the AAV vectors or virus of the invention for the treatment of a disease or disorder.

[0025] In yet another embodiment, the disease or disorder in a subject is associated with abnormal activity of an endogenous protein. As used herein "endogenous protein" means a protein or gene product encoded by the genome of the subject suffering from the disease or disorder.

[0026] An "AAV virion" or "AAV viral particle" or "AAV vector particle" or "AAV virus" refers to a viral particle composed of at least one AAV capsid protein and an encapsidated polynucleotide AAV vector. If the particle comprises a heterologous polynucleotide (i.e. a polynucleotide other than a wild-type AAV genome such as a transgene to be delivered to a mammalian cell), it is typically referred to as an "AAV vector particle" or simply an "AAV

vector". Thus, production of AAV vector particle necessarily includes production of AAV vector, as such a vector is contained within an AAV vector particle.

[0027] The invention also provides for cells comprising any of the AAV vectors of the invention, and viral particles produced by these cells of the invention.

[0028] The term "inverted terminal repeat (ITR)" as used herein refers to the art-recognized regions found at the 5' and 3' termini of the AAV genome which function in cis as origins of DNA replication and as packaging signals for the viral genome. AAV ITRs, together with the AAV rep coding region, provide for efficient excision and rescue from a plasmid vector, and integration of a nucleotide sequence interposed between two flanking ITRs into a host cell genome. Sequences of certain AAV-associated ITRs are disclosed by Yan et al., *J. Virol.* 79(1):364-379 (2005) which is herein incorporated by reference in its entirety.

[0029] The phrase "helper functions for generating a productive AAV infection" as used herein refers to AAV-derived coding sequences that can be expressed to provide AAV gene products that, in turn, function in trans for productive AAV replication. Thus, AAV helper functions include the rep and cap regions. The rep expression products have been shown to possess many functions, including, among others: recognition, binding and nicking of the AAV origin of DNA replication; DNA helicase activity; and modulation of transcription from AAV (or other heterologous) promoters. The cap expression products supply necessary packaging functions. AAV helper functions are used herein to complement AAV functions in trans that are missing from AAV vectors. Helper functions for generating a productive AAV infection also may include certain helper functions from baculovirus, herpes virus, adenovirus, or vaccinia virus.

[0030] In some embodiments, the viral construct comprises a nucleotide sequence encoding AAV rep and cap genes.

[0031] The term "AAV rep gene" as used herein refers to the art-recognized region of the AAV genome which encodes the replication proteins of the virus which are required to replicate the viral genome and to insert the viral genome into a host genome during latent infection. For a further description of the AAV rep coding region, see, e.g., Muzyczka et al., *Current Topics in Microbiol. and Immunol.* 158:97-129 (1992); Kotin et al., *Human Gene Therapy* 5:793-801 (1994), the disclosures of which are incorporated herein by reference in their entireties. The rep

coding region, as used herein, can be derived from any viral serotype, such as the AAV serotypes described above. The region need not include all of the wild-type genes but may be altered, e.g., by the insertion, deletion or substitution of nucleotides, so long as the rep genes retain the desired functional characteristics when expressed in a suitable recipient cell.

[0032] The term "AAV cap gene" as used herein refers to the art-recognized region of the AAV genome which encodes the coat proteins of the virus which are required for packaging the viral genome. For a further description of the cap coding region, see, e.g., Muzyczka et al., *Current Topics in Microbiol. and Immunol.* 158:97-129 (1992); Kotin et al., *Human Gene Therapy* 5:793-801 (1994), the disclosures of which are incorporated herein by reference in their entireties. The AAV cap coding region, as used herein, can be derived from any AAV serotype, as described above. The region need not include all of the wild-type cap genes but may be altered, e.g., by the insertion, deletion or substitution of nucleotides, so long as the genes provide for sufficient packaging functions when present in a host cell along with an AAV vector.

[0033] The term "transfection" is used to refer to the uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al., *Virology* 52:456 (1973); Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratories, New York (1989); Davis et al., Basic Methods in Molecular Biology, Elsevier (1986); Chu et al., *Gene* 13:197 (1981), the disclosures of which are incorporated herein by reference in their entireties. Such techniques can be used to introduce one or more exogenous DNA moieties, such as a nucleotide integration vector and other nucleic acid molecules, into suitable host cells. The term captures chemical, electrical, and viral-mediated transfection procedures.

[0034] In yet another aspect, described herein is an AAV particle produced by a method described herein. In some embodiments, the AAV particle comprises in its genome at least one nucleotide encoding a heterologous protein.

[0035] The term "heterologous proteins or peptides" refer to any protein that is not expressed by wild type AAV including tags such as hexahistidine, FLAG, myc, polyhistidine, or labels or immunogens, adjuvants, selection markers, therapeutic proteins or targeting proteins or peptides, to name a few.

[0036] Exemplary heterologous protein described herein includes, but is not limited to, β-globin, hemoglobin, tissue plasminogen activator, and coagulation factors; colony stimulating factors (CSF); interleukins, such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, etc.; growth factors, such as keratinocyte growth factor (KGF), stem cell factor (SCF), fibroblast growth factor (FGF, such as basic FGF and acidic FGF), hepatocyte growth factor (HGF), insulin-like growth factors (IGFs), bone morphogenetic protein (BMP), epidermal growth factor (EGF), growth differentiation factor-9 (GDF-9), hepatoma derived growth factor (HDGF), myostatin (GDF-8), nerve growth factor (NGF), neurotrophins, platelet-derived growth factor (PDGF), thrombopoietin (TPO), transforming growth factor alpha (TGF-α), transforming growth factor beta (TGF-β), and the like; soluble receptors, such as soluble TNF-α receptors, soluble interleukin receptors (e.g., soluble IL-1 receptors and soluble type II IL-1 receptors), soluble γ/Δ T cell receptors, ligand-binding fragments of a soluble receptor, and the like; enzymes, such as α-glucosidase, imiglucerase, β-glucocerebrosidase, and alglucerase; enzyme activators, such as tissue plasminogen activator; chemokines, such as 1P-10, monokine induced by interferon-gamma (Mig), Groα/IL-8, RANTES, MIP-1α, MIP-1β, MCP-1, PF-4, and the like; angiogenic agents, such as vascular endothelial growth factors (VEGFs, e.g., VEGF121, VEGF165, VEGF-C, VEGF-2), glioma-derived growth factor, angiogenin, angiogenin-2; and the like; anti-angiogenic agents, such as a soluble VEGF receptor; protein vaccine; neuroactive peptides, such as nerve growth factor (NGF), bradykinin, cholecystokinin, gastin, secretin, oxytocin, gonadotropin-releasing hormone, beta-endorphin, enkephalin, substance P, somatostatin, prolactin, galanin, growth hormone-releasing hormone, bombesin, dynorphin, warfarin, neurotensin, motilin, thyrotropin, neuropeptide Y, luteinizing hormone, calcitonin, insulin, glucagons, vasopressin, angiotensin II, thyrotropin-releasing hormone, vasoactive intestinal peptide, a sleep peptide, and the like; thrombolytic agents; atrial natriuretic peptide; relaxin; glial fibrillary acidic protein; follicle stimulating hormone (FSH); human alpha-1 antitrypsin; leukemia inhibitory factor (LIF); tissue factors, luteinizing hormone; macrophage activating factors; tumor necrosis factor (TNF); neutrophil chemotactic factor (NCF); tissue inhibitors of metalloproteinases; vasoactive intestinal peptide; angiogenin; angiotropin; fibrin; hirudin; IL-1 receptor antagonists; ciliary neurotrophic factor (CNTF); brain-derived neurotrophic factor (BDNF); neurotrophins 3 and 4/5 (NT-3 and 4/5); glial cell derived neurotrophic factor (GDNF); aromatic amino acid decarboxylase (AADC); Factor VIII, Factor IX, Factor X; dystrophin or

nini-dystrophin; lysosomal acid lipase; phenylalanine hydroxylase (PAH); glycogen storage disease-related enzymes, such as glucose-6-phosphatase, acid maltase, glycogen debranching enzyme, muscle glycogen phosphorylase, liver glycogen phosphorylase, muscle phosphofructokinase, phosphorylase kinase, glucose transporter, aldolase A, β-enolase, glycogen synthase; and lysosomal enzymes.

DESCRIPTION OF DRAWINGS

[0037] Figure 1 is a set of tables showing the relative tissue specificity of infectivity of AAV having the designated novel capsid proteins. The data is presented as Total Flux in each tissue (photons/sec/cm^2/radian). For each tissue the top row represents the average flux and the bottom row represents the standard deviation.

[0038] Figure 2 provides data from an IVIG neutralization assay for the novel capsids: Bba-45, Bba-46, Bba-47, Bba-50 and Bba-51 and control capsids: AAV5, AAV8 and AAV9. This assay demonstrates that the novel capsids exhibited IVIG resistant properties.

[0039] Figure 3 provides transduction data for the novel capsids: Bba-45, Bba-46, Bba-47, Bba-50 and Bba-51 and control capsids: AAV5, AAV8 and AAV9. This data was generated in multiple experiments and the multiple data points for each capsid represents different experiments. The AAV having the novel capsids have a significantly higher degree of specificity for liver cells compared to AAV5. The data is shown as total flux activity, which is a proxy for AAV infectivity of each organ system.

[0040] Figure 4 provides transduction data in multiple tissues for the novel capsids: Bba-45, Bba-46, Bba-47, Bba-50 and Bba-51 and control capsids: AAV5, AAV8 and AAV9. This data demonstrates that the AAV having the novel capsids have a significantly higher degree of specificity for liver cells compared to AAV5.

[0041] Figure 5 provides the plasma bCG protein levels after injection of the novel capsids: Bba-45, Bba-46, Bba-47, Bba-49 and Bba-50 and control capsids: AAV8, AAV-rh10 and AAV-anc80L65.

[0042] Figure 6 provides the level of bCG protein levels in the liver and plasma after injection of the novel capsids: Bba-45, Bba-46, Bba-47, Bba-49 and Bba-50 and control capsids: AAV8, AAV-rh10 and AAV-anc80L65.

[0043] Figures 7A-7C provide the level of bCG DNA (Panel A), bCG RNA (Panel B) and bCG protein (Panel C) in the liver after injection of the novel capsids: Bba-45, Bba-46, Bba-47, Bba-49 and Bba-50 and control capsids: AAV8, AAV-rh10 and AAV-anc80L65.

[0044] Figure 8A-8B demonstrates that the bCG DNA and bCG RNA levels in the liver correlate with the bCG protein levels in the liver after injection of the novel capsids: Bba-45, Bba-46, Bba-47, Bba-49 and Bba-50 and control capsids: AAV8, AAV-rh10 and AAV-anc80L65.

[0045] Figure 9 provides the percentage of hepatocytes transduced by AAV after injection of the novel capsids: Bba-45, Bba-46, Bba-47, Bba-49 and Bba-50 and control capsids: AAV8, AAV-rh10 and AAV-anc80L65 as determined by immunohistochemistry.

[0046] Figure 10 provides exemplary photos showing mmunohistochemistry staining of hepatocytes after injection of the novel capsids: Bba-45, Bba-46, Bba-47, Bba-49 and Bba-50 and control capsids: AAV8, AAV-rh10 and AAV-anc80L65.

[0047] Figure 11 provides the percentage of hepatocytes transduced by AAV having the novel capsid Bba-49 and the AAV5 capsid from various sources. These AAV transduced the hepatocytes with the –AGXT transgene. The AAV.Bba-49. AGXT transduced about 96% of the hepatocytes.

DETAILED DESCRIPTION

The invention provides for novel AAV capsid proteins, nucleic acid encoding those capsid proteins and AAV virus comprising those novel capsid proteins. In some embodiments, the AAV capsid proteins were isolated and identified from various mammalian tissues. The amino acid sequences of the novel AAV capsid VP1 proteins are set out as SEQ ID NOS:1-7, and the locations of the associated VP2 and VP3 regions are disclosed herein.

AAV Vectors

[0048] As used herein, the term "AAV" is a standard abbreviation for adeno-associated virus. Adeno-associated virus is a single-stranded DNA parvovirus that grows only in cells in which certain functions are provided by a co-infecting helper virus. There are currently at least thirteen serotypes of AAV that have been characterized, as shown below in Table 1. General information and reviews of AAV can be found in, for example, Carter, Handbook of Parvoviruses, Vol. 1, pp.

169-228 (1989), and Berns, Virology, pp. 1743-1764, Raven Press, (New York, 1990). However, it is fully expected that these same principles will be applicable to additional AAV serotypes since it is well known that the various serotypes are quite closely related, both structurally and functionally, even at the genetic level. (See, for example, Blacklowe, pp. 165-174 of Parvoviruses and Human Disease, J. R. Pattison, ed. (1988); and Rose, *Comprehensive Virology* 3:1-61 (1974)). For example, all AAV serotypes apparently exhibit very similar replication properties mediated by homologous rep genes; and all bear three related capsid proteins such as those expressed in AAV6. The degree of relatedness is further suggested by heteroduplex analysis which reveals extensive cross-hybridization between serotypes along the length of the genome; and the presence of analogous self-annealing segments at the termini that correspond to "inverted terminal repeat sequences" (ITRs). The similar infectivity patterns also suggest that the replication functions in each serotype are under similar regulatory control.

[0049] An "AAV vector" as used herein refers to a vector comprising one or more polynucleotides of interest (or transgenes) that are flanked by AAV terminal repeat sequences (ITRs). Such AAV vectors can be replicated and packaged into infectious viral particles when present in a host cell that has been transfected with a vector encoding and expressing rep and cap gene products.

[0050] An "AAV virion" or "AAV viral particle" or "AAV vector particle" or "AAV virus" refers to a viral particle composed of at least one AAV capsid protein and an encapsidated polynucleotide AAV vector. If the particle comprises a heterologous polynucleotide (i.e. a polynucleotide other than a wild-type AAV genome such as a transgene to be delivered to a mammalian cell), it is typically referred to as an "AAV vector particle" or simply an "AAV vector". Thus, production of AAV vector particle necessarily includes production of AAV vector, as such a vector is contained within an AAV vector particle.

[0051] AAV "rep" and "cap" genes are genes encoding replication and encapsidation proteins, respectively. AAV rep and cap genes have been found in all AAV serotypes examined to date, and are described herein and in the references cited. In wild-type AAV, the rep and cap genes are generally found adjacent to each other in the viral genome (i.e., they are "coupled" together as adjoining or overlapping transcriptional units), and they are generally conserved among AAV serotypes. AAV rep and cap genes are also individually and collectively referred to as "AAV

packaging genes." The AAV cap gene in accordance with the present invention encodes a Cap protein which is capable of packaging AAV vectors in the presence of rep and adeno helper function and is capable of binding target cellular receptors. In some embodiments, the AAV cap gene encodes a capsid protein having an amino acid sequence derived from a particular AAV serotype, for example the serotypes shown in Table 1.

Table 1. AAV serotypes

AAV Serotype	Genbank Accession No.
AAV-1	NC_002077.1
AAV-2	NC_001401.2
AAV-3	NC_001729.1
AAV-3B	AF028705.1
AAV-4	NC_001829.1
AAV-5	NC_006152.1
AAV-6	AF028704.1
AAV-7	NC_006260.1
AAV-8	NC_006261.1
AAV-9	AX753250.1
AAV-10	AY631965.1
AAV-11	AY631966.1
AAV-12	DQ813647.1
AAV-13	EU285562.1

[0052] The AAV sequences employed for the production of AAV can be derived from the genome of any AAV serotype. Generally, the AAV serotypes have genomic sequences of significant homology at the amino acid and the nucleic acid levels, provide a similar set of genetic functions, produce virions which are essentially physically and functionally equivalent, and replicate and assemble by practically identical mechanisms. For the genomic sequence of AAV serotypes and a discussion of the genomic similarities see, for example, GenBank Accession number U89790; GenBank Accession number J01901; GenBank Accession number AF043303; GenBank Accession number AF085716; Chlorini et al., *J. Vir.* 71:6823-33(1997); Srivastava et al., *J. Vir.* 45:555-64 (1983); Chlorini et al., *J. Vir.* 73:1309-1319 (1999); Rutledge et al., *J. Vir.* 72:309-319 (1998); and Wu et al., *J. Vir.* 74: 8635-47 (2000).

[0053] The genomic organization of all known AAV serotypes is very similar. The genome of AAV is a linear, single-stranded DNA molecule that is less than about 5,000 nucleotides (nt) in length. Inverted terminal repeats (ITRs) flank the unique coding nucleotide sequences for the non-structural replication (Rep) proteins and the structural (VP) proteins. The VP proteins form the capsid. The terminal 145 nt are self-complementary and are organized so that an energetically stable intramolecular duplex forming a T-shaped hairpin may be formed. These hairpin structures function as an origin for viral DNA replication, serving as primers for the cellular DNA polymerase complex. The Rep genes encode the Rep proteins, Rep78, Rep68,

Rep52, and Rep40. Rep78 and Rep68 are transcribed from the p5 promoter, and Rep 52 and Rep40 are transcribed from the p19 promoter. The cap genes encode the VP proteins, VP1, VP2, and VP3. The cap genes are transcribed from the p40 promoter.

[0054] In some embodiments, a nucleic acid sequence encoding an AAV capsid protein is operably linked to regulatory expression control sequences for expression in a specific cell type, such as Sf9 or HEK cells. Techniques known to one skilled in the art for expressing foreign genes in insect host cells or mammalian host cells can be used to practice the invention. Methodology for molecular engineering and expression of polypeptides in insect cells is described, for example, in Summers and Smith. A Manual of Methods for Baculovirus Vectors and Insect Culture Procedures, Texas Agricultural Experimental Station Bull. No. 7555, College Station, Tex. (1986); Luckow. 1991. In Prokop et al., Cloning and Expression of Heterologous Genes in Insect Cells with Baculovirus Vectors' Recombinant DNA Technology and Applications, 97-152 (1986); King, L. A. and R. D. Possee, The baculovirus expression system, Chapman and Hall, United Kingdom (1992); O'Reilly, D. R., L. K. Miller, V. A. Luckow, Baculovirus Expression Vectors: A Laboratory Manual, New York (1992); W.H. Freeman and Richardson, C. D., Baculovirus Expression Protocols, Methods in Molecular Biology, volume 39 (1995); U.S. Pat. No. 4,745,051; US2003148506; and WO 03/074714. A particularly suitable promoter for transcription of a nucleotide sequence encoding an AAV capsid protein is e.g. the polyhedron promoter. However, other promoters that are active in insect cells are known in the art, e.g. the p10, p35 or IE-1 promoters and further promoters described in the above references are also contemplated.

[0055] Use of insect cells for expression of heterologous proteins is well documented, as are methods of introducing nucleic acids, such as vectors, e.g., insect-cell compatible vectors, into such cells and methods of maintaining such cells in culture. See, for example, METHODS IN MOLECULAR BIOLOGY, ed. Richard, Humana Press, NJ (1995); O'Reilly et al., BACULOVIRUS EXPRESSION VECTORS, A LABORATORY MANUAL, Oxford Univ. Press (1994); Samulski et al., *J. Vir.* 63:3822-8 (1989); Kajigaya et al., *Proc. Nat'l. Acad. Sci. USA* 88:4646-50 (1991); Ruffing et al., *J. Vir.* 66:6922-30 (1992); Kirnbauer et al., *Vir.* 219:37-44 (1996); Zhao et al., *Vir.* 272:382-93 (2000); and Samulski et al., U.S. Pat. No. 6,204,059. In some embodiments, the nucleic acid construct encoding AAV in insect cells is an insect cell-compatible vector. An "insect cell-compatible vector" or "vector" as used herein refers to a

nucleic acid molecule capable of productive transformation or transfection of an insect or insect cell. Exemplary biological vectors include plasmids, linear nucleic acid molecules, and recombinant viruses. Any vector can be employed as long as it is insect cell-compatible. The vector may integrate into the insect cell's genome but the presence of the vector in the insect cell need not be permanent and transient episomal vectors are also included. The vectors can be introduced by any means known, for example by chemical treatment of the cells, electroporation, or infection. In some embodiments, the vector is a baculovirus, a viral vector, or a plasmid. In a more preferred embodiment, the vector is a baculovirus, i.e. the construct is a baculoviral vector. Baculoviral vectors and methods for their use are described in the above cited references on molecular engineering of insect cells.

[0056] Baculoviruses are enveloped DNA viruses of arthropods, two members of which are well known expression vectors for producing recombinant proteins in cell cultures. Baculoviruses have circular double-stranded genomes (80-200 kbp) which can be engineered to allow the delivery of large genomic content to specific cells. The viruses used as a vector are generally *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) or Bombyx mori (Bm)NPV) (Kato et al., *Appl. Microbiol. Biotechnol.* 85(3):459-470 (2010)). Baculoviruses are commonly used for the infection of insect cells for the expression of recombinant proteins. In particular, expression of heterologous genes in insects can be accomplished as described in for instance U.S. Pat. No. 4,745,051; Friesen et al., *Curr. Top. Microbiol. Immunol.* 131:31-49. (1986); EP 127,839; EP 155,476; Miller et al., *Ann. Rev. of Microbiol.* 42: 177-199 (1988); Carbonell et al., *Gene* 73(2):409-18 (1988); Maeda et al., *Nature* 315(6020):592-4 (1985); Lebacq-Verheyden et al., *Mol. Cell. Biol.* 8(8):3129-35 (1988); Smith et al., *Proc. Natl. Acad. Sci. U S A.* 82(24):8404-8 (1985); Miyajima et al., *Gene* 58(2-3):273-81 (1987); and Martin et al., *DNA* 7(2):99-106 (1988). Numerous baculovirus strains and variants and corresponding permissive insect host cells that can be used for protein production are described in Luckow et al., *Nature Biotechnology* 6:47-55 (1988), and Maeda et al., *Nature* 315(6020):592-4 (1985).

Novel AAV Capsid Protein

[0057] In a first aspect, the invention provides for novel AAV capsid proteins that were isolated from various mammalian tissues. The novel AAV VP1 capsid proteins are provided as SEQ ID NOS:1-7 and the locations of the associated VP2 and VP3 regions are described herein.

The invention also provides for polynucleotides comprising a nucleotide sequence encoding these novel AAV capsid proteins. The invention provides the amino acid sequences of the novel AAV capsid proteins (referred herein collectively as the "AAV capsid proteins of the invention"), and the nucleic acid sequences encoding the AAV capsid proteins of the invention. Also provided are fragments of these AAV capsid nucleic acid and amino acid sequences of the invention. Each of these sequences may be readily utilized in a variety of vector systems and host cells. Desirable fragments of the capsid VP1 proteins include VP2, VP3 and variable regions, the GBS domain and the GH loop, and polynucleotide sequences encoding these proteins. These fragments may be readily utilized in a variety of vector systems and host cells. Such fragments may be used alone, in combination with other AAV sequences or fragments, or in combination with elements from other AAV or non-AAV viral sequences. In one particularly desirable embodiment, a vector contains the AAV capsid sequences of the invention.

[0058] The AAV capsid sequences of the invention and fragments thereof are useful in production of rAAV, and are also useful as antisense delivery vectors, gene therapy vectors, or vaccine vectors. The invention further provides nucleic acid molecules, gene delivery vectors, and host cells which contain the novel AAV capsid sequences of the invention.

[0059] Suitable fragments can be determined using the information provided herein. Alignments are performed using any of a variety of publicly or commercially available Multiple Sequence Alignment Programs, such as "Clustal W", accessible through Web Servers on the internet. Alternatively, Vector NTI utilities are also used. There are also a number of algorithms known in the art which can be used to measure nucleotide sequence identity, including those contained in the programs described above. As another example, polynucleotide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) as provided in GCG Version 6.1, herein incorporated by reference. Similar programs are available for amino acid sequences, e.g., the "Clustal X" program. Additional sequence alignment tools that can be used are provided by (protein sequence alignment; (http://www.ebi.ac.uk/Tools/psa/emboss_needle/)) and (nucleic acid alignment; http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)). Generally, any of these

programs are used at default settings, although one of skill in the art can alter these settings as needed. Alternatively, one of skill in the art can utilize another algorithm or computer program which provides at least the level of identity or alignment as that provided by the referenced algorithms and programs.

[0060] The terms “substantial identity”, "substantial homology" or "substantial similarity," when referring to a nucleic acid, or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 95 to 99% of the aligned sequences such as 95% identity, 96% identity, 97% identity, 98% identity and 99% identity. Preferably, the homology is over the full-length of the two sequences being compared, or an open reading frame thereof, or another suitable fragment which is at least 15 nucleotides in length. Examples of suitable fragments are described herein. Also included in the nucleic acid sequences of the invention are natural variants and engineered modifications of the nucleic acids encoding the AAV capsids of the invention and its complementary strand. Such modifications include, for example, labels which are known in the art, methylation, and substitution of one or more of the naturally occurring nucleotides with a degenerate nucleotide.

[0061] The terms “substantial identity”, "substantial homology" or "substantial similarity," when referring to amino acids or fragments thereof, indicates that, when optimally aligned with appropriate amino acid insertions or deletions with another amino acid (or its complementary strand), there is amino acid sequence identity in at least about 95 to 99% of the aligned sequences such as 95% identity, 96% identity, 97% identity, 98% identity and 99% identity. Preferably, the homology is over the full-length of the two sequences being compared, or a protein thereof, e.g., a cap protein, a rep protein, or a fragment thereof which is at least 8 amino acids, or more desirably, at least 15 amino acids in length. Examples of suitable fragments are described herein.

[0062] By the term "highly conserved" is meant at least 80% identity, preferably at least 90% identity, and more preferably, over 97% identity. Identity is readily determined by one of skill in the art by resort to algorithms and computer programs known by those of skill in the art.

[0063] The term "percent sequence identity" or "identical" in the context of nucleic acid sequences or amino acid sequences refers to the residues in the two sequences which are the

same when aligned for maximum correspondence. The length of sequence identity comparison may be over the full-length of the two sequences being compared, the full-length of a gene coding sequence, or a fragment of at least about 500 to 5000 nucleotides, is desired. However, identity among smaller fragments, e.g. of at least about nine nucleotides, usually at least about 20 to 24 nucleotides, at least about 28 to 32 nucleotides, at least about 36 or more nucleotides, may also be desired. Similarly, "percent sequence identity" may be readily determined for amino acid sequences, over the full-length of a protein, or a fragment thereof. Suitably, a fragment is at least about 8 amino acids in length, and may be up to about 700 amino acids. Examples of suitable fragments are described herein.

[0064] The novel capsids of the invention may comprise one or more additional conservative amino acid substitutions that do not affect the biological and/or immunogenic activity of the polypeptide. The term "conservative amino acid substitution" refers to a substitution of a native amino acid residue with a nonnative residue, including naturally occurring and nonnaturally occurring amino acids, such that there is little or no effect on the polarity or charge of the amino acid residue at that position. For example, a conservative substitution results from the replacement of a non-polar residue in a polypeptide with any other non-polar residue. Further, any native residue in the polypeptide may also be substituted with alanine, according to the methods of "alanine scanning mutagenesis". Naturally occurring amino acids are characterized based on their side chains as follows: basic: arginine, lysine, histidine; acidic: glutamic acid, aspartic acid; uncharged polar: glutamine, asparagine, serine, threonine, tyrosine; and non-polar: phenylalanine, tryptophan, cysteine, glycine, alanine, valine, proline, methionine, leucine, norleucine, isoleucine. General rules for amino acid substitutions are set forth in the Table below.

Conservative Amino Acid Substitutions

Original Residues	Exemplary Substitutions	Preferred Substitutions
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln	Gln
Asp	Glu	Glu
Cys	Ser, Ala	Ser
Gln	Asn	Asn

Glu	Asp	Asn
Gly	Pro, Ala	Ala
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe,	Leu
Leu	Norleucine, Ile, Val, Met,	Leu
Lys	Arg, 1,4 Diaminobutyric	Arg
Met	Leu, Phe, Ile	Leu
Phe	Leu, Val, Ile, Ala, Tyr	Arg
Pro	Ala	Gly
Ser	Thr, Ala, Cys	Thr
Thr	Ser	Ser
Trp	Tyr, Phe	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Met, Leu, Phe, Ala,	Leu

[0065] The novel capsids of the invention may be encoded by polynucleotides that are substantially equivalent to the nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 1-7. Polynucleotides according to the invention can have, e.g., at least 65%, at least 70%, at least 75%, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more typically at least 90%, 91%, 92%, 93%, or 94% and even more typically at least 95%, 96%, 97%, 98% or 99% sequence identity to the polynucleotide sequences encoding the modified polypeptide amino acid sequences of the invention.

[0066] Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to the nucleotide sequences encoding the novel capsids of the invention, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g., 15, 17, or 20 nucleotides or more that are selective for (i.e., specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or

can differentiate genes from other bacterial genes, and are preferably based on unique nucleotide sequences.

[0067] The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Hybridization stringency is principally determined by temperature, ionic strength, and the concentration of denaturing agents such as formamide. Examples of stringent conditions for hybridization and washing are 0.015 M sodium chloride, 0.0015M sodium citrate at 65-68°C or 0.015 M sodium chloride, 0.0015M sodium citrate, and 50% formamide at 42 °C See Sambrook et al., Molecular Cloning: A Laboratory Manual, 2.sup.nd Ed., Cold Spring Harbor Laboratory, (Cold Spring Harbor, N.Y. 1989). More stringent conditions (such as higher temperature, lower ionic strength, higher formamide, or other denaturing agent) may also be used, however, the rate of hybridization will be affected. In instances wherein hybridization of deoxyoligonucleotides is concerned, additional exemplary stringent hybridization conditions include washing in 6xSSC 0.05% sodium pyrophosphate at 37 °C (for 14-base oligos), 48 °C (for 17-base oligos), 55 °C (for 20-base oligos), and 60 °C. (for 23-base oligos).

[0068] Other agents may be included in the hybridization and washing buffers for the purpose of reducing non-specific and/or background hybridization. Examples are 0.1% bovine serum albumin, 0.1% polyvinyl-pyrrolidone, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate, NaDodSO4 (SDS), ficoll, Denhardt's solution, sonicated salmon sperm DNA (or other non-complementary DNA), and dextran sulfate, although other suitable agents can also be used. The concentration and types of these additives can be changed without substantially affecting the stringency of the hybridization conditions. Hybridization experiments are usually carried out at pH 6.8-7.4, however, at typical ionic strength conditions, the rate of hybridization is nearly independent of pH. See Anderson et al., Nucleic Acid Hybridisation: A Practical Approach, Ch. 4, IRL Press Limited (Oxford, England). Hybridization conditions can be adjusted by one skilled in the art in order to accommodate these variables and allow DNAs of different sequence relatedness to form hybrids.

[0069] As described herein, the vectors of the invention containing or comprising the AAV capsid proteins of the invention are particularly well-suited for use in applications in which the neutralizing antibodies diminish the effectiveness of other AAV serotype based vectors, as well

as other viral vectors. The rAAV vectors of the invention are particularly advantageous in rAAV re-administration and repeat gene therapy.

[0070] Also included within the invention are fragments of the nucleic acids encoding the AAV capsid proteins of the invention, their complementary strand, cDNA and RNA complementary thereto. Suitable fragments are at least 15 nucleotides in length, and encompass functional fragments, i.e., fragments which are of biological interest. Such fragments include the sequences encoding the three variable proteins (VP) of the capsid which are alternative splice variants: VP1, VP2 and VP3. Other suitable fragments of the nucleic acids encoding the AAV capsids of the invention include the fragment which contains the start codon for the capsid protein, and the fragments encoding the variable regions of the VP1 capsid protein, which are described herein.

[0071] The invention is not limited to the AAV capsid amino acid sequences, peptides and proteins expressed from the AAV nucleic acid sequences of the invention and encompasses amino acid sequences, peptides and proteins generated by other methods known in the art, including, e.g., by chemical synthesis, by other synthetic techniques, or by other methods. For example, the sequences of any of the capsids described herein can be readily generated using a variety of techniques.

[0072] Suitable production techniques are well known to those of skill in the art. See, e.g., Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, N.Y.). Alternatively, peptides can also be synthesized by the well-known solid phase peptide synthesis methods (Merrifield, *J. Am. Chem. Soc.*, 85:2149 (1962); Stewart and Young, Solid Phase Peptide Synthesis Freeman, (San Francisco, 1969) pp. 27-62. These and other suitable production methods are within the knowledge of those of skill in the art and are not a limitation of the present invention.

[0073] The AAV capsid is composed of three proteins, VP1, VP2 and VP3, which are alternative splice variants. The full-length capsid sequence is referred to as VP1 which encompasses the spliced variants referred to as VP2 and VP3. The invention also provides for other functional fragments of the AAV capsid proteins of the invention. Other desirable fragments of the capsid protein include the variable regions (VR), the constant regions which are

located between the variable regions, the GBS domain, and the GH loop. Other desirable fragments of the capsid protein include the HPV themselves.

[0074] An algorithm has been developed to determine areas of sequence divergence in AAV2. (Chiorini et al, *J. Virol*, 73:1309-19 (1999); Rutledge et al, *J. Virol.*, 72:309-319 (1998)). Using this algorithm and/or the alignment techniques described herein, the VR of the novel AAV capsid sequences are determined. Using the alignment provided herein performed using the Clustal X program at default settings, or using other commercially or publicly available alignment programs at default settings, one of skill in the art can readily determine corresponding fragments of the novel AAV capsids of the invention.

[0075] Suitably, fragments of an AAV capsid protein are at least 8 amino acids in length, or at least 9 amino acids in length, or at least 10 amino acids in length, or least 20 amino acids in length, or 30 amino acids in length or at least 50 amino acids in length, or at least 75 amino acids in length, or at least 100 amino acids in length or 200 amino acids in length or 250 amino acids in length or 300 amino acids in length or 350 amino acids in length or 400 amino acids in length. However, fragments of other desired lengths may be readily utilized. All fragments of the invention retain biological activity of a capsid AAV protein. Such fragments may be produced recombinantly or by other suitable means, e.g., chemical synthesis.

[0076] The sequences, proteins, and fragments of the invention may be produced by any suitable means, including recombinant production, chemical synthesis, or other synthetic means. Such production methods are within the knowledge of those of skill in the art and are not a limitation of the present invention.

[0077] In addition to including the nucleic acid sequences provided in the figures and Sequence Listing, the present invention includes nucleic acid molecules and sequences which are designed to express the amino acid sequences, proteins and peptides of the AAV capsid proteins of the invention. Thus, the invention includes nucleic acid sequences which encode the following AAV capsid amino acid sequences and artificial AAV capsid proteins generated using these sequences and/or unique fragments thereof.

Production of AAV with the Capsid Proteins of the Invention

[0078] The invention encompasses AAV capsid protein sequences and the nucleic acids encoding these proteins of which are free of DNA and/or cellular material which these viruses are associated in nature. In another aspect, the present invention provides molecules which utilize the novel AAV sequences of the invention, including fragments thereof, for production of molecules useful in delivery of a heterologous gene or other nucleic acid sequences to a target cell.

[0079] In another aspect, the present invention provides molecules which utilize the AAV capsid protein sequences of the invention, including fragments thereof, for production of viral vectors useful in delivery of a heterologous gene or other nucleic acid sequences to a target cell.

[0080] The molecules of the invention which contain AAV capsid nucleic acid sequences include any genetic element (vector) which may be delivered to a host cell, e.g., naked DNA, a plasmid, phage, transposon, cosmid, episome, a protein in a non-viral delivery vehicle (e.g., a lipid-based carrier), virus, etc. which transfer the sequences carried thereon. The selected vector may be delivered by any suitable method, including transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion. The methods used to construct any embodiment of this invention are known to those with skill in nucleic acid manipulation and include genetic engineering, recombinant engineering, and synthetic techniques. See, e.g., Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

[0081] In one embodiment, the vectors of the invention contain, at a minimum, sequences encoding the AAV capsid of the invention or a fragment thereof. In another embodiment, the vectors of the invention contain, at a minimum, sequences encoding an AAV rep protein or a fragment thereof. Optionally, such vectors may contain both AAV cap and rep proteins. In vectors in which both AAV rep and cap are provided, the AAV rep and AAV cap sequences can both be of the same AAV serotype origin. Alternatively, the present invention provides vectors in which the rep sequences are from an AAV serotype which differs from that which is providing the cap sequences. In one embodiment, the rep and cap sequences are expressed from separate sources (e.g., separate vectors, or a host cell and a vector). In another embodiment, these rep

sequences are fused in frame to cap sequences of a different AAV serotype to form a chimeric AAV vector.

[0082] Thus, in one embodiment, the vectors described herein contain nucleic acid sequences encoding an intact AAV capsid protein of any one of amino acid sequences SEQ ID NO: 1-7. In another example, it may be desirable to alter the start codon of the VP3 protein to GTG. Alternatively, the rAAV may contain one or more of the variable regions of one or more of the AAV capsid proteins of the invention, or other fragments. These modifications may be to increase expression, yield, and/or to improve purification in the selected expression systems, or for another desired purpose (e.g., to change tropism or alter neutralizing antibody epitopes).

[0083] The vectors described herein, e.g., a plasmid, are useful for a variety of purposes, but are particularly well suited for use in production of a rAAV containing a capsid comprising AAV sequences or a fragment thereof. These vectors, including rAAV, their elements, construction, and uses are described in detail herein.

VP1 of Novel Capsid Proteins

[0084] Novel AAV VP1 capsid proteins were isolated from baboon liver.

[0085] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.45) is set out as SEQ ID NO:1 (amino acids 1-742) and the locations of the associated variable regions and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans amino acids 138-742 of SEQ ID NO:1 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:1.

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANRQHQDNARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQQRLATDTSFGGNLGKAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRALDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGYFDFNRFHCHFSPRDWQRLINNNWGLRPKSMRVKIFNIQVREVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPFHSMYAHSQ
SLDRMMNPLLDQYLWHLQSTTTGNSLNQGTATTTYGKITTGDFAYYRKNWLPGACIKQQKFSKNASQNYKIPASGGDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNATTAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0086] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.46) is set out as SEQ ID NO:2 (amino acids 1-742) and the locations of the associated variable regions

and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans amino acids 138-742 of SEQ ID NO:2 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:2.

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQQRLATDTSFGGNLGKAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGCFDFNRFHCHFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPFHSMYAHSQ
SLDRMMNPLLDQYLWHLQSTTTGNSLNQGAATTTYGKITTGDFAYYRKNWLPGACIKQQKFSKNASQNYKIPASGGDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNATTAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0087] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.47) is set out as SEQ ID NO:3 (amino acids 1-742) and the locations of the associated variable regions and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans amino acids 138-742 of SEQ ID NO:3 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:3.

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQQRLATDTSFGGNLGKAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGYFDFNRFHCHFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPFHSMYAHSQ
SLDRMMNPLLDQYLWHLQSTTTGNSLNQGTATTTYGKITTGDFAYYRKNWLPGACIKQQKFSKNASQNYKIPASGGDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNATTAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0088] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.48) is set out as SEQ ID NO:4 (amino acids 1-742) and the locations of the associated variable regions and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans amino acids 138-742 of SEQ ID NO:4 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:4.

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQERLQEDTSFGGNLGRAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGYFDFNRFHCRFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPFHSMYAHSQ

SLDRMMNPLLDQYLWHLQSTTTGNSLNQGTAITTYGKITTGDFAYYRKNWLPGACIKQQKFSKNASQNYKIPASGGDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNAATAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0089] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.49) is set out as SEQ ID NO:5 (amino acids 1-742) and the locations of the associated variable regions and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans amino acids 138-742 of SEQ ID NO:5 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:5.

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQERLQEDTSFGGNLGRAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGYFDFNRFHCHFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPFHSMYAHSQ
SLDRMMNPLLDQYLWHLQSTTTGNSLNQGTAITTYGKITTGDFAYYRKNWLPGAGIKQQKFSKNASQNYKIPASGGDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNATTAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0090] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.50) is set out as SEQ ID NO:6 (amino acids 1-742) and the locations of the associated variable regions and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans amino acids 138-742 of SEQ ID NO:6 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:6.

MAADGYLPDWLEDNLSESIREWWALKPGAPRPKANQQHQDDARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQQRLATDTSFGGNLGKAVFQAKKRILEPLGLVEEGVKTAPGRKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGYFDFNRFHCHFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPLHSMYAHSQ
SLDRMMNPLLDQYLWHLQSTTTGNSLNQGTATTTYGKITTGDFAYYRKNWLPGACIKQQKFSKNASQNYKIPASGEDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNATTAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0091] The VP1 sequence of a novel AAV capsid isolated from baboon (denoted as Bba.51) is set out as SEQ ID NO:7 (amino acids 1-742) and the locations of the associated variable regions and GBS and GH loop regions are defined in Table 2 below. The VP2 capsid protein spans

amino acids 138-742 of SEQ ID NO:7 and the VP3 capsid protein spans amino acids 206-742 of SEQ ID NO:7.

MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYLGPGNGLDKGEPVNEADAAALEHDKAYDQQLKSG
DNPYLKYNHADAEFQQRLATDTSFGGNLGKAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPAKKKQKDGET
ADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGGNAVEAGQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPTYNNH
LYLRIGTTANSNTYNGFSTPWGYFDFNRFHCHFSPRDWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSNGETTVANNLTSTVQIFA
DSTYELPYVMDAGQEGSLPPFPNDVFMVPQYGYCGVVTGENQNQTDRNAFYCLEYFPSQMLRTGNNFEISYQFEKVPFHSMYAHSQ
SLDRMMNPLLDQYLWHLQSTTTGNSLNQGTATTTYGKITTGDFAYYRKNWLPGACIKQQKFSKNASQNYKIPASGGDALLKYDTHT
TLNGRWSNMAPGPPMATAGAGDSDFSNSQLIFAGPNQSGNTTTSSNNLLFTSEEEIATTNPRDTDMFGQIADNNQNATTAPHIANL
DAMGIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTPVPANPNTTFSAARINSFLTQYSTGQVAVQ
IDWEIQKEHSKRWNPEVQFTSNYGTQNSMLWAPDNAGNYHEPRAIGSRFLTHHL

[0092] The corresponding nucleic acid sequences encoding the above referenced capsid proteins are SEQ ID NO:8 / Bba.45; SEQ ID NO:9 / Bba.46; SEQ ID NO:10 / Bba.47; SEQ ID NO:11 / Bba.48; SEQ ID NO:12 / Bba.49; SEQ ID NO:13 / Bba.50; and SEQ ID NO:14 /Bba.51.

[0093] SEQ ID NO:8 (Bba.45)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACAGCCCAAGGCAAATCGACAACATCAAGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTCCAGCAGCGCTTGGCGACCGACACCTCTTTTGGGGGCAACCTCGG
CAAGGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAAGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGGCACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG
ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATACTTTGACTTTAACCGCTT
CCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAGGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTTTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AACAGCTACCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTACCTGGAGCCTGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATTCCCGCCAGCGGGGGAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCACCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGGGCCAAGGTCCCTCACAC

GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[0094] SEQ ID NO:9 (Bba.46)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACAGCCCAAGGCAAATCAACAACATCAAGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTCCAGCAGCGCTTGGCGACCGACACCTCTTTTGGGGGCAACCTCGG
CAAGGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAAGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGACGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG
ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATGCTTTGACTTTAACCGCTT
CCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAAGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTTTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AGCAGCTACCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTGCCTGGAGCCTGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATCCCCGCCAGCGGGGGAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCACCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGGGCCAAGGTCCCTCACAC
GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[0095] SEQ ID NO:10 (Bba.47)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACAGCCCAAGGCAAATCAACAACATCAAGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTCCAGCAGCGCTTGGCGACCGACACCTCTTTTGGGGGCAACCTCGG
CAAGGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAAGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCTCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG

ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATACTTTGACTTTAACCGCTT
CCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAAGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTTTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AACAGCTACCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTGCCTGGAGCCTGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATTCCCGCCAGCGGGGGAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGGCAGATTGCAGATAATAATCAAAATGCCACCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGGGCCAAGGTCCCTCACAC
GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[0096] SEQ ID NO:11 (Bba.48)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACAGCCCAAGGCAAATCAACAACATCAAGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTTCAGGAGCGTCTTCAAGAAGATACGTCTTTTGGGGGCAACCTCGG
GCGAGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAAGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG
ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATACTTTGACTTTAACCGCTT
CCACTGCCGCTTTTCCCCGCGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAAGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTTTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AACAGCTATCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTGCCTGGAGCCTGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATTCCCGCCAGCGGGGGAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCGCCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGGGCCAAGGTCCCTCACAC

GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[0097] SEQ ID NO:12 (Bba.49)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACAGCCCAAGGCAAATCAACAACATCAAGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTTCAGGAGCGTCTTCAAGAAGATACGTCTTTTGGGGGCAACCTCGG
GCGAGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAAGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG
ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATACTTTGACTTTAACCGCTT
CCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAAGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTTTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AACAGCTATCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTGCCTGGAGCCGGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATTCCCGCCAGCGGGGGAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCACCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGGGCCAAGGTCCCTCACAC
GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[0098] SEQ ID NO:13 (Bba.50)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAAGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACGGCCCAAGGCAAATCAACAACATCAAGACGACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTCCAGCAGCGCTTGGCGACCGACACCTCTTTTGGGGGCAACCTCGG
CAAGGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAGGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG

ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATACTTTGACTTTAACCGCTT
CCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAAGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTCTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AACAGCTACCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTGCCTGGAGCCTGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATTCCCGCCAGCGGGGAAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCACCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATCTGGGCCAAGGTCCCTCACAC
GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[0099] SEQ ID NO:14 (Bba.51)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTCGCGAGTGGTGGGCGCTGAAACCTGGAGC
CCCACAGCCCAAGGCAAATCAACAACATCAAGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAATACTTGGGACCCGGTAACG
GACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGCTCAAGTCGGGA
GACAACCCGTACCTCAAGTACAACCACGCGGACGCCGAGTTCCAGCAGCGCTTGGCGACCGACACCTCTTTTGGGGGCAACCTCGG
CAAGGCAGTCTTCCAGGCCAAAAAGAGGATTCTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCCTGGAAAGAAAC
GCCCATTAGAAAAGACTCCAAATCGGCCGACCAACCCGGACTCTGGGAAGGCCCCGGCCAAGAAAAAGCAAAAAGACGGCGAGACA
GCCGACTCTGCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCCTGAGGGATCATCTTCCGGAGAAATGTC
TCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTCGAGGCGGGACAAGGTGCCGATGGAGTGGGTAATGCCTCCGGTG
ATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGGGTCCTGCCCACCTACAACAACCAC
CTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAATGGATTCTCCACCCCCTGGGGATACTTTGACTTTAACCGCTT
CCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATCAACAACAACTGGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCT
TCAACATCCAGGTCAAGGAGGTCACTACGTCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCG
GATTCAACGTATGAACTCCCATACGTGATGGACGCCGGTCAGGAGGGGAGCCTTCCTCCGTTCCCCAACGACGTGTTTATGGTTCC
CCAATACGGGTACTGCGGAGTCGTCACTGGAGAAAACCAGAACCAAACAGACAGAAATGCCTTTTACTGTCTGGAGTACTTTCCAT
CCCAAATGCTAAGAACTGGCAACAACTTTGAAATCAGTTACCAATTTGAAAAAGTTCCTTTCCATTCAATGTACGCGCACAGCCAG
AGCCTGGACAGAATGATGAATCCTTTGCTGGATCAGTACCTGTGGCATCTGCAATCGACCACTACCGGAAATTCCCTTAATCAAGG
AACAGCTACCACCACGTACGGGAAAATTACCACTGGGGACTTTGCCTACTACAGGAAAAACTGGTTGCCTGGAGCCTGCATTAAAC
AACAAAAATTTTCAAAGAATGCCAGTCAAAACTACAAGATTCCCGCCAGCGGGGGAGACGCCCTTTTAAAGTATGACACGCATACC
ACTTTAAATGGGCGATGGAGTAACATGGCTCCTGGTCCTCCAATGGCCACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCA
GCTGATCTTTGCCGGACCCAATCAGAGCGGTAACACGACCACGTCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCA
CAACAAACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCACCACCGCCCCTCACATCGCTAACCTG
GACGCTATGGGAATTGTTCCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGGGCCAAGGTCCCTCACAC

GGACGGACACTTTCACCCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCTCCGCAGATTTTCATCAAAAACACCCCCG
TACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATCAATTCTTTTTTGACGCAGTACAGCACCGGACAAGTCGCCGTTCAG
ATCGACTGGGAAATTCAGAAGGAGCACTCCAAACGCTGGAATCCCGAAGTCCAATTTACTTCAAACTACGGCACTCAAAATTCTAT
GCTGTGGGCTCCCGACAACGCCGGCAACTACCACGAACCCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

[00100] In Table 2 immediately below, "VR" refers to the variable region and the numbers refer to the amino acid residues each variable region or the GBS and GH loop regions span in the amino acid sequence.

Table 2

AAV	VRI	VRII	VRIII	VRIV	GBS	VRV	VRVI	VRVII
AAVBba.45	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564
AAVBba.46	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564
AAVBba.47	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564
AAVBba.48	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564
AAVBba.49	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564
AAVBba.50	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564
AAVBba.51	265-269	325-330	381-386	452-462	468-478	490-512	534-544	550-564

Transgene

[00101] The transgene is a nucleic acid sequence, heterologous to the vector sequences flanking the transgene, which encodes a polypeptide, protein, or other product, of interest. The nucleic acid coding sequence is operatively linked to regulatory components in a manner which permits transgene transcription, translation, and/or expression in a host cell.

[00102] The composition of the transgene sequence will depend upon the use to which the resulting vector will be put. For example, one type of transgene sequence includes a reporter sequence, which upon expression produces a detectable signal. Such reporter sequences include, without limitation, DNA sequences encoding β-lactamase, β-galactosidase (LacZ), alkaline phosphatase, thymidine kinase, green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), luciferase, membrane bound proteins including, for example, CD2, CD4, CD8, the influenza hemagglutinin protein, and others well known in the art, to which high

affinity antibodies directed thereto exist or can be produced by conventional means, and fusion proteins comprising a membrane bound protein appropriately fused to an antigen tag domain from, among others, hemagglutinin or Myc.

[00103] These coding sequences, when associated with regulatory elements which drive their expression, provide signals detectable by conventional means, including enzymatic, radiographic, colorimetric, fluorescence or other spectrographic assays, fluorescent activating cell sorting assays and immunological assays, including enzyme linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and immunohistochemistry. For example, where the marker sequence is the LacZ gene, the presence of the vector carrying the signal is detected by assays for beta-galactosidase activity. Where the transgene is green fluorescent protein or luciferase, the vector carrying the signal may be measured visually by color or light production in a luminometer.

[00104] However, desirably, the transgene is a non-marker sequence encoding a product which is useful in biology and medicine, such as proteins, peptides, RNA, enzymes, dominant negative mutants, or catalytic RNAs. Desirable RNA molecules include tRNA, dsRNA, ribosomal RNA, catalytic RNAs, siRNA, small hairpin RNA, trans-splicing RNA, and antisense RNAs. One example of a useful RNA sequence is a sequence which inhibits or extinguishes expression of a targeted nucleic acid sequence in the treated animal. Typically, suitable target sequences include oncologic targets and viral diseases. See, for examples of such targets the oncologic targets and viruses identified below in the section relating to immunogens.

[00105] The transgene may be used to correct or ameliorate gene deficiencies, which may include deficiencies in which normal genes are expressed at less than normal levels or deficiencies in which the functional gene product is not expressed. A preferred type of transgene sequence encodes a therapeutic protein or polypeptide which is expressed in a host cell. The invention further includes using multiple transgenes, e.g., to correct or ameliorate a gene defect caused by a multi-subunit protein. In certain situations, a different transgene may be used to encode each subunit of a protein, or to encode different peptides or proteins. This is desirable when the size of the DNA encoding the protein subunit is large, e.g., for an immunoglobulin, the platelet-derived growth factor, or a dystrophin protein. In order for the cell to produce the multi-subunit protein, a cell is infected with the recombinant virus containing each of the different

subunits. Alternatively, different subunits of a protein may be encoded by the same transgene. In this case, a single transgene includes the DNA encoding each of the subunits, with the DNA for each subunit separated by an internal ribozyme entry site (IRES). This is desirable when the size of the DNA encoding each of the subunits is small, e.g., the total size of the DNA encoding the subunits and the IRES is less than five kilobases. As an alternative to an IRES, the DNA may be separated by sequences encoding a 2A peptide, which self-cleaves in a post-translational event. See, e.g., Donnelly *et al, J. Gen. Virol.*, 78(Pt 1):13-21 (January 1997); Furler, *et al, Gene Ther.*, 8(11):864-873 (June 2001); Klump *et al., Gene Ther.*, 8(10):811-817 (May 2001). This 2A peptide is significantly smaller than an IRES, making it well suited for use when space is a limiting factor. More often, when the transgene is large, consists of multi-subunits, or two transgenes are co-delivered, rAAV carrying the desired transgene(s) or subunits are co-administered to allow them to concatamerize *in vivo* to form a single vector genome. In such an embodiment, a first AAV may carry an expression cassette which expresses a single transgene and a second AAV may carry an expression cassette which expresses a different transgene for co-expression in the host cell. However, the selected transgene may encode any biologically active product or other product, e.g., a product desirable for study.

[00106] Suitable transgenes may be readily selected by one of skill in the art. The selection of the transgene is not considered to be a limitation of this invention.

[00107] In some embodiments, the transgene is a heterologous protein, and this heterologous protein is a therapeutic protein. Exemplary therapeutic proteins include, but are not limited to, blood factors, such as β-globin, hemoglobin, tissue plasminogen activator, and coagulation factors; colony stimulating factors (CSF); interleukins, such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, etc.; growth factors, such as keratinocyte growth factor (KGF), stem cell factor (SCF), fibroblast growth factor (FGF, such as basic FGF and acidic FGF), hepatocyte growth factor (HGF), insulin-like growth factors (IGFs), bone morphogenetic protein (BMP), epidermal growth factor (EGF), growth differentiation factor-9 (GDF-9), hepatoma derived growth factor (HDGF), myostatin (GDF-8), nerve growth factor (NGF), neurotrophins, platelet-derived growth factor (PDGF), thrombopoietin (TPO), transforming growth factor alpha (TGF-α.), transforming growth factor beta (TGF-.β.), and the like; soluble receptors, such as soluble TNF-α. receptors, soluble VEGF receptors, soluble interleukin receptors (e.g., soluble IL-1 receptors and soluble type II IL-1 receptors), soluble .γ/δ T cell receptors, ligand-binding fragments of a soluble

receptor, and the like; enzymes, such as α-glucosidase, imiglucarase, β-glucocerebrosidase, and alglucerase; enzyme activators, such as tissue plasminogen activator; chemokines, such as 1P-10, monokine induced by interferon-gamma (Mig), Groα/IL-8, RANTES, MIP-1α, MIP-1β., MCP-1, PF-4, and the like; angiogenic agents, such as vascular endothelial growth factors (VEGFs, e.g., VEGF121, VEGF165, VEGF-C, VEGF-2), glioma-derived growth factor, angiogenin, angiogenin-2; and the like; anti-angiogenic agents, such as a soluble VEGF receptor; protein vaccine; neuroactive peptides, such as nerve growth factor (NGF), bradykinin, cholecystokinin, gastin, secretin, oxytocin, gonadotropin-releasing hormone, beta-endorphin, enkephalin, substance P, somatostatin, prolactin, galanin, growth hormone-releasing hormone, bombesin, dynorphin, warfarin, neurotensin, motilin, thyrotropin, neuropeptide Y, luteinizing hormone, calcitonin, insulin, glucagons, vasopressin, angiotensin II, thyrotropin-releasing hormone, vasoactive intestinal peptide, a sleep peptide, and the like; thrombolytic agents; atrial natriuretic peptide; relaxin; glial fibrillary acidic protein; follicle stimulating hormone (FSH); human alpha-1 antitrypsin; leukemia inhibitory factor (LIF); tissue factors, luteinizing hormone; macrophage activating factors; tumor necrosis factor (TNF); neutrophil chemotactic factor (NCF); tissue inhibitors of metalloproteinases; vasoactive intestinal peptide; angiogenin; angiotropin; fibrin; hirudin; IL-1 receptor antagonists; and the like. Some other non-limiting examples of protein of interest include ciliary neurotrophic factor (CNTF); brain-derived neurotrophic factor (BDNF); neurotrophins 3 and 4/5 (NT-3 and 4/5); glial cell derived neurotrophic factor (GDNF); aromatic amino acid decarboxylase (AADC); hemophilia related clotting proteins, such as Factor VIII, Factor IX, Factor X; dystrophin, mini-dystrophin, or microdystrophin; lysosomal acid lipase; phenylalanine hydroxylase (PAH); glycogen storage disease-related enzymes, such as glucose-6-phosphatase, acid maltase, glycogen debranching enzyme, muscle glycogen phosphorylase, liver glycogen phosphorylase, muscle phosphofructokinase, phosphorylase kinase (e.g., PHKA2), glucose transporter (e.g., GLUT2), aldolase A, β-enolase, and glycogen synthase; lysosomal enzymes (e.g., beta-N-acetylhexosaminidase A); and any variants thereof.

Regulatory Control Elements

[00108] The AAV vector also includes conventional control elements or sequences which are operably linked to the transgene in a manner which permits its transcription, translation and/or expression in a cell transfected with the plasmid vector or infected with the virus produced by the invention. As used herein, "operably linked" sequences include both expression control

sequences that are contiguous with the gene of interest and expression control sequences that act in trans or at a distance to control the gene of interest.

[00109] Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation (polyA) signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*i.e.*, Kozak consensus sequence); sequences that enhance protein stability; and when desired, sequences that enhance secretion of the encoded product. A great number of expression control sequences, including promoters which are native, constitutive, inducible and/or tissue-specific, are known in the art and may be utilized.

[00110] Examples of constitutive promoters include, without limitation, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) (*see,* e.g., Boshart *et al, Cell*, 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 promoter [Invitrogen]. Inducible promoters allow regulation of gene expression and can be regulated by exogenously supplied compounds, environmental factors such as temperature, or the presence of a specific physiological state, e.g., acute phase, a particular differentiation state of the cell, or in replicating cells only. Inducible promoters and inducible systems are available from a variety of commercial sources, including, without limitation, Invitrogen, Clontech and Ariad. Many other systems have been described and can be readily selected by one of skill in the art. Examples of inducible promoters regulated by exogenously supplied compounds, include, the zinc-inducible sheep metallothionine (MT) promoter, the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter, the T7 polymerase promoter system [WO 98/10088]; the ecdysone insect promoter [No *et al, Proc. Natl. Acad. Sci. USA*, 93:3346-3351 (1996)], the tetracycline-repressible system [Gossen *et al, Proc. Natl. Acad. Sci. USA*, 89:5547-5551 (1992)], the tetracycline-inducible system [Gossen *et al, Science*, 268:1766-1769 (1995), see also Harvey *et al, Curr. Opin. Chem. Biol.*, 2:512-518 (1998)], the RU486-inducible system [Wang *et al, Nat. Biotech.*, 15:239-243 (1997) and Wang *et al, Gene Ther.*, 4:432-441 (1997)] and the rapamycin-inducible system [Magari *et al, J. Clin. Invest.*, 100:2865-2872 (1997)]. Other types of inducible promoters which may be useful in this context are those which are regulated by a specific

physiological state, e.g., temperature, acute phase, a particular differentiation state of the cell, or in replicating cells only.

[00111] In another embodiment, the native promoter for the transgene will be used. The native promoter may be preferred when it is desired that expression of the transgene should mimic the native expression. The native promoter may be used when expression of the transgene must be regulated temporally or developmentally, or in a tissue-specific manner, or in response to specific transcriptional stimuli. In a further embodiment, other native expression control elements, such as enhancer elements, polyadenylation sites or Kozak consensus sequences may also be used to mimic the native expression.

[00112] Another embodiment of the transgene includes a gene operably linked to a tissue-specific promoter. For instance, if expression in skeletal muscle is desired, a promoter active in muscle should be used. These include the promoters from genes encoding skeletal β-actin, myosin light chain 2A, dystrophin, muscle creatine kinase, as well as synthetic muscle promoters with activities higher than naturally-occurring promoters (see *Li et al., Nat. Biotech.*, 17:241-245 (1999)). Examples of promoters that are tissue-specific are known for liver (albumin, Miyatake *et al., J. Virol.*, 71:5124-32 (1997); hepatitis B virus core promoter, Sandig *et al., Gene Ther.*, 3:1002-9 (1996); alpha-fetoprotein (AFP), Arbuthnot *et al., Hum. Gene Ther.*, 7:1503-14 (1996)), bone osteocalcin (Stein *et al., Mol. Biol. Rep.*, 24:185-96 (1997)); bone sialoprotein (Chen *et al., J. Bone Miner. Res.*, 11:654-64 (1996)), lymphocytes (CD2, Hansal e*t al., J. Immunol.*, 161:1063-8 (1998); immunoglobulin heavy chain; T cell receptor chain), neuronal such as neuron-specific enolase (NSE) promoter (Andersen *et al., Cell. Mol. Neurobiol.*, 13:503-15 (1993)), neurofilament light-chain gene (Piccioli *et al., Proc. Natl. Acad. Sci. USA*, 88:5611-5 (1991)), and the neuron-specific vgf gene (Piccioli *et al., Neuron*, 15:373-84 (1995)), among others.

[00113] Optionally, plasmids carrying therapeutically useful transgenes may also include selectable markers or reporter genes may include sequences encoding geneticin, hygromicin or purimycin resistance, among others. Such selectable reporters or marker genes (preferably located outside the viral genome to be rescued by the method of the invention) can be used to signal the presence of the plasmids in bacterial cells, such as ampicillin resistance. Other components of the plasmid may include an origin of replication. Selection of these and other

promoters and vector elements are conventional and many such sequences are available [see, e.g., Sambrook et al, and references cited therein].

Methods for Producing Recombinant AAVs

[00114] The present disclosure provides materials and methods for producing recombinant AAVs in insect or mammalian cells. In some embodiments, the viral construct further comprises a promoter and a restriction site downstream of the promoter to allow insertion of a polynucleotide encoding one or more proteins of interest, wherein the promoter and the restriction site are located downstream of the 5' AAV ITR and upstream of the 3' AAV ITR. In some embodiments, the viral construct further comprises a posttranscriptional regulatory element downstream of the restriction site and upstream of the 3' AAV ITR. In some embodiments, the viral construct further comprises a polynucleotide inserted at the restriction site and operably linked with the promoter, where the polynucleotide comprises the coding region of a protein of interest. As a skilled artisan will appreciate, any one of the AAV vector disclosed in the present application can be used in the method as the viral construct to produce the recombinant AAV.

[00115] In some embodiments, the helper functions are provided by one or more helper plasmids or helper viruses comprising adenoviral or baculoviral helper genes. Non-limiting examples of the adenoviral or baculoviral helper genes include, but are not limited to, E1A, E1B, E2A, E4 and VA, which can provide helper functions to AAV packaging.

[00116] Helper viruses of AAV are known in the art and include, for example, viruses from the family *Adenoviridae* and the family *Herpesviridae*. Examples of helper viruses of AAV include, but are not limited to, SAdV-13 helper virus and SAdV-13-like helper virus described in US Publication No. 20110201088 (the disclosure of which is incorporated herein by reference), helper vectors pHELP (Applied Viromics). A skilled artisan will appreciate that any helper virus or helper plasmid of AAV that can provide adequate helper function to AAV can be used herein.

[00117] In some embodiments, the AAV cap genes are present in a plasmid. The plasmid can further comprise an AAV rep gene. The cap genes and/or rep gene from any AAV serotype (including, but not limited to, AAV1, AAV2, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV13 and any variants thereof) can be used herein to produce the recombinant AAV. In some embodiments, the AAV cap genes encode a capsid from serotype 1,

serotype 2, serotype 4, serotype 5, serotype 6, serotype 7, serotype 8, serotype 9, serotype 10, serotype 11, serotype 12, serotype 13 or a variant thereof.

[00118] In some embodiments, the insect or mammalian cell can be transfected with the helper plasmid or helper virus, the viral construct and the plasmid encoding the AAV cap genes; and the recombinant AAV virus can be collected at various time points after co-transfection. For example, the recombinant AAV virus can be collected at about 12 hours, about 24 hours, about 36 hours, about 48 hours, about 72 hours, about 96 hours, about 120 hours, or a time between any of these two time points after the co-transfection.

[00119] Recombinant AAV can also be produced using any conventional methods known in the art suitable for producing infectious recombinant AAV. In some instances, a recombinant AAV can be produced by using an insect or mammalian cell that stably expresses some of the necessary components for AAV particle production. For example, a plasmid (or multiple plasmids) comprising AAV rep and cap genes, and a selectable marker, such as a neomycin resistance gene, can be integrated into the genome of the cell. The insect or mammalian cell can then be co-infected with a helper virus (e.g., adenovirus or baculovirus providing the helper functions) and the viral vector comprising the 5' and 3' AAV ITR (and the nucleotide sequence encoding the heterologous protein, if desired). The advantages of this method are that the cells are selectable and are suitable for large-scale production of the recombinant AAV. As another non-limiting example, adenovirus or baculovirus rather than plasmids can be used to introduce rep and cap genes into packaging cells. As yet another non-limiting example, both the viral vector containing the 5' and 3' AAV LTRs and the rep-cap genes can be stably integrated into the DNA of producer cells, and the helper functions can be provided by a wild-type adenovirus to produce the recombinant AAV.

Cell Types Used in AAV Production

[00120] The viral particles comprising the AAV vectors of the invention may be produced using any invertebrate cell type which allows for production of AAV or biologic products and which can be maintained in culture. For example, the insect cell line used can be from *Spodoptera frugiperda*, such as Sf9, SF21, SF900+, drosophila cell lines, mosquito cell lines, e.g., *Aedes albopictus* derived cell lines, domestic silkworm cell lines, e.g. Bombyxmori cell lines, *Trichoplusia ni* cell lines such as High Five cells or Lepidoptera cell lines such as

Ascalapha odorata cell lines. Preferred insect cells are cells from the insect species which are susceptible to baculovirus infection, including High Five, Sf9, Se301, SeIZD2109, SeUCR1, Sf900+, Sf21, BTI-TN-5B1-4, MG-1, Tn368, HzAm1, BM-N, Ha2302, Hz2E5 and Ao38.

[00121] Baculoviruses are enveloped DNA viruses of arthropods, two members of which are well known expression vectors for producing recombinant proteins in cell cultures. Baculoviruses have circular double-stranded genomes (80-200 kbp) which can be engineered to allow the delivery of large genomic content to specific cells. The viruses used as a vector are generally *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) or *Bombyx mori* (Bm-NPV) (Kato et al., 2010).

[00122] Baculoviruses are commonly used for the infection of insect cells for the expression of recombinant proteins. In particular, expression of heterologous genes in insects can be accomplished as described in for instance U.S. Pat. No. 4,745,051; Friesen et al (1986); EP 127,839; EP 155,476; Vlak et al. (1988); Miller et al. (1988); Carbonell et al. (1988); Maeda et al. (1985); Lebacq-Verheyden et al. (1988); Smith et al. (1985); Miyajima et al. (1987); and Martin et al.(1988). Numerous baculovirus strains and variants and corresponding permissive insect host cells that can be used for protein production are described in Luckow et al (1988), Miller et al. (1986); Maeda et al. (1985) and McKenna (1989).

[00123] In another aspect of the invention, the methods of the invention are also carried out with any mammalian cell type which allows for replication of AAV or production of biologic products, and which can be maintained in culture. Preferred mammalian cells used can be HEK293, HeLa, CHO, NS0, SP2/0, PER.C6, Vero, RD, BHK, HT 1080, A549, Cos-7, ARPE-19 and MRC-5 cells.

Production of Heterologous Proteins In Vitro

[00124] As a non-limiting example, the recombinant AAV disclosed herein can be used to produce a protein of interest *in vitro*, for example, in a cell culture. As one non-limiting example, in some embodiments, a method for producing a protein of interest *in vitro*, where the method includes providing a recombinant AAV comprising a nucleotide sequence encoding the heterologous protein; and contacting the recombinant AAV with a cell in a cell culture, whereby the recombinant AAV expresses the protein of interest in the cell. The size of the nucleotide sequence encoding the protein of interest can vary. For example, the nucleotide sequence can be

at least about 1.4 kb, at least about 1.5 kb, at least about 1.6 kb, at least about 1.7 kb, at least about 1.8 kb, at least about 2.0 kb, at least about 2.2 kb, at least about 2.4 kb, at least about 2.6 kb, at least about 2.8 kb, at least about 3.0 kb, at least about 3.2 kb, at least about 3.4 kb, at least about 3.5 kb in length, at least about 4.0 kb in length, at least about 5.0 kb in length, at least about 6.0 kb in length, at least about 7.0 kb in length, at least about 8.0 kb in length, at least about 9.0 kb in length, or at least about 10.0 kb in length. In some embodiments, the nucleotide is at least about 1.4 kb in length.

Production of Heterologous Proteins In Vivo

[00125] The recombinant AAV disclosed herein can be used to produce a protein of interest *in vivo*, for example in an animal such as a mammal. Some embodiments provide a method for producing a protein of interest *in vivo*, where the method includes providing a recombinant AAV comprising a nucleotide sequence encoding the protein of interest; and administering the recombinant AAV to the subject, whereby the recombinant AAV expresses the protein of interest in the subject. The subject can be, in some embodiments, a non-human mammal, for example, a monkey, a dog, a cat, a mouse, or a cow. The size of the nucleotide sequence encoding the protein of interest can vary. For example, the nucleotide sequence can be at least about 1.4 kb, at least about 1.5 kb, at least about 1.6 kb, at least about 1.7 kb, at least about 1.8 kb, at least about 2.0 kb, at least about 2.2 kb, at least about 2.4 kb, at least about 2.6 kb, at least about 2.8 kb, at least about 3.0 kb, at least about 3.2 kb, at least about 3.4 kb, at least about 3.5 kb in length, at least about 4.0 kb in length, at least about 5.0 kb in length, at least about 6.0 kb in length, at least about 7.0 kb in length, at least about 8.0 kb in length, at least about 9.0 kb in length, or at least about 10.0 kb in length. In some embodiments, the nucleotide is at least about 1.4 kb in length.

Therapeutic Uses

[00126] The recombinant AAV produced by the methods described can be used to express one or more therapeutic proteins to treat various diseases or disorders. Non-limiting examples of the diseases include cancer such as carcinoma, sarcoma, leukemia, lymphoma; and autoimmune diseases such as multiple sclerosis. Non-limiting examples of carcinomas include esophageal carcinoma; hepatocellular carcinoma; basal cell carcinoma, squamous cell carcinoma (various tissues); bladder carcinoma, including transitional cell carcinoma; bronchogenic carcinoma; colon carcinoma; colorectal carcinoma; gastric carcinoma; lung carcinoma, including small cell

carcinoma and non-small cell carcinoma of the lung; adrenocortical carcinoma; thyroid carcinoma; pancreatic carcinoma; breast carcinoma; ovarian carcinoma; prostate carcinoma; adenocarcinoma; sweat gland carcinoma; sebaceous gland carcinoma; papillary carcinoma; papillary adenocarcinoma; cystadenocarcinoma; medullary carcinoma; renal cell carcinoma; ductal carcinoma in situ or bile duct carcinoma; choriocarcinoma; seminoma; embryonal carcinoma; Wilm's tumor; cervical carcinoma; uterine carcinoma; testicular carcinoma; osteogenic carcinoma; epithelieal carcinoma; and nasopharyngeal carcinoma. Non-limiting examples of sarcomas include fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, chordoma, osteogenic sarcoma, osteosarcoma, angiosarcoma, endotheliosarcoma, lymphangiosarcoma, lymphangioendotheliosarcoma, synovioma, mesothelioma, Ewing's sarcoma, leiomyosarcoma, rhabdomyosarcoma, and other soft tissue sarcomas. Non-limiting examples of solid tumors include glioma, astrocytoma, medulloblastoma, craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendroglioma, menangioma, melanoma, neuroblastoma, and retinoblastoma. Non-limiting examples of leukemias include chronic myeloproliferative syndromes; acute myelogenous leukemias; chronic lymphocytic leukemias, including B-cell CLL, T-cell CLL prolymphocytic leukemia, and hairy cell leukemia; and acute lymphoblastic leukemias. Examples of lymphomas include, but are not limited to, B-cell lymphomas, such as Burkitt's lymphoma; Hodgkin's lymphoma; and the like. Other non-liming examples of the diseases that can be treated using the AAV vectors, recombinant viruses and methods disclosed herein include genetic disorders including sickle cell anemia, cystic fibrosis, lysosomal acid lipase (LAL) deficiency 1, Tay-Sachs disease, Phenylketonuria, Mucopolysaccharidoses, Glycogen storage diseases (GSD, e.g., GSD types I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, and XIV), Galactosemia, muscular dystrophy (e.g., Duchenne muscular dystrophy), and hemophilia such as hemophilia A (classic hemophilia) and hemophilia B (Christmas Disease), Wilson's disease, Fabry Disease, Gaucher Disease hereditary angioedema (HAE), and alpha 1 antitrypsin deficiency. In addition, the AAV vectors, recombinant viruses and methods disclosed herein can be used to other disorders that can be treated by local expression of a transgene in the liver or by expression of a secreted protein from the liver or a hepatocyte.

[00127] The amount of the heterologous protein expressed in the subject (e.g., the serum of the subject) can vary. For example, in some embodiments the protein can be expressed in the

serum of the subject in the amount of at least about 9 μg/ml, at least about 10 μg/ml, at least about 50 μg/ml, at least about 100 μg/ml, at least about 200 μg/ml, at least about 300 μg/ml, at least about 400 μg/ml, at least about 500 μg/ml, at least about 600 μg/ml, at least about 700 μg/ml, at least about 800 μg/ml, at least about 900 μg/ml, or at least about 1000 μg/ml. In some embodiments, the protein of interest is expressed in the serum of the subject in the amount of about 9 μg/ml, about 10 μg/ml, about 50 μg/ml, about 100 μg/ml, about 200 μg/ml, about 300 μg/ml, about 400 μg/ml, about 500 μg/ml, about 600 μg/ml, about 700 μg/ml, about 800 μg/ml, about 900 μg/ml, about 1000 μg/ml, about 1500 μg/ml, about 2000 μg/ml, about 2500 μg/ml, or a range between any two of these values. A skilled artisan will understand that the expression level in which a protein of interest is needed for the method to be effective can vary depending on non-limiting factors such as the particular protein of interest and the subject receiving the treatment, and an effective amount of the protein can be readily determined by a skilled artisan using conventional methods known in the art without undue experimentation.

EXAMPLES

Example 1

Isolation of Novel Naturally-Occurring Capsid Proteins

[00128] Novel naturally-occurring capsid proteins were isolated from baboon liver. Frozen liver tissue was obtained from Texas Biomedical or the New England Primate Research Center. Genomic DNA was prepared from liver tissue using the DNeasy Blood & Tissue kit (Qiagen catalog #69504).

[00129] Polymerase chain reaction (PCR) was carried out on the genomic DNA using the following primers: primer rep-1397-F (5'-GTGCCCTTTTACGGCTGCGTGAACTGGACCAATGAAAACTTTCC- 3'SEQ ID NO:21) and primer cap-2872-R (5'-CCGACGGAGTGGGCAATGCCTCAGGAAATTGGCATTG CGATTCC- 3' SEQ ID NO:22) under the following conditions: initial incubation: 97°C, 120 sec, denaturation step: 97°C, 15 sec, annealing step: 58°C, 60°C, or 62°C, 15 sec, extension step: 72°C, 240 sec. The denaturation, annealing, and extension steps were performed for 35 cycles. Then the reaction was incubated at 72°C, 7 min and stored at 4°C until analyzed. The

PCR products were separated by electrophoresis on 1% agarose gels, isolated using the Gel Extraction Kit (Qiagen catalog #28704), and cloned into pCR4-TOPO-TA (Invitrogen catalog # 450030) according to the manufacturer's instructions. After transformation of *E. coli*, NEB5α cells, DNA was prepared from ampicillin resistant colonies and sequenced from both ends to determine if the insert encoded an AAV-related sequence.

[00130] If the inserts in pCR4-TOPO TA were related to AAV sequences, sequence-specific primers were designed to the rep portion of the sequence to perform "around the episome PCR" (hereinafter "ATE PCR") to obtain a complete capsid gene. ATE PCR is based on the notion that persistent AAV genomes forms circular episomes in animal tissues. Accordingly, one can use a "divergent" set of primers corresponding to a sequence in the rep gene to perform polymerase chain reactions to isolate most or all of any AAV sequence that may exist in that episome but in particular one could isolate a complete contiguous capsid gene. Multimers of episomes can form, for example by homologous recombination, and in that case it is possible to isolate more than one capsid gene (which usually are not the same) from a single ATE PCR reaction.

[00131] An ATE PCR was carried-out in a standard polymerase chain reaction instrument using a 2-step program as follows: initial incubation: 95°C, 240 sec, denaturation step: 95°C, 30 sec annealing / extension step: 72°C, 300 sec. The denaturation and combined annealing / extension steps were performed for 40 cycles. The reaction was then incubated at 72°C, 7 min and stored at 4°C until analyzed. The PCR products were electrophoresed on 1% agarose gels. PCR products that were the length of multimers of an AAV genome (~4.5 kilobases) were excised from the gel, purified using the QIAquick Gel Extraction Kit (Qiagen catalog #28704), and cloned into pCR4-TOPO-TA (Invitrogen catalog # 450030) according to the manufacturer's instructions. After transformation of *E. coli*, NEB5α cells, DNA was prepared from ampicillin resistant colonies and the entire sequence of the insert was determined.

[00132] If the 2-step program described above did not produce PCR products of the correct size the following 3-step program was used: Initial incubation: 95°C, 240 sec, Denaturation step: 95°C, 30 sec Annealing step: 62°C, 64°C, 66°C, or 68°C, 30 sec, Extension step: 72°C, 300 sec. The denaturation, annealing, and extension steps were performed for 40 cycles. Then the reaction was incubated at 72°C, 7 min and stored at 4°C until analyzed as above.

[00133] Once complete insert sequences in pCR4-TOPO TA were determined they were identified as being AAV capsid genes using the BLAST algorithm (available at the NCBI website). Their relationship to known AAVs was determined using various nucleotide or amino acid sequence alignment programs such as Clustal Omega (available at the EBI web site) or Vector NTI (Invitrogen, Inc.).

[00134] To produce AAV, the unique AAV capsid genes were subcloned into an expression plasmid (pAAV-RC; Agilent, Inc.), then transfected into 293 cells along with a vector (pAAV luciferase) and adenovirus helper plasmid (pHELPER; Agilent, Inc.). AAV production was allowed to occur for 3 days and then crude lysates were made by freeze-thawing the cells three times. Debris was pelleted and the supernatant (crude AAV) was titered by Q-PCR to determine a genomic titer (which confirms the capsid is capable of assembly and DNA packaging) and then used to assess transduction by the AAVs on various cells.

[00135] The VP1 amino acid sequences of the novel mammalian tissue-derived AAV capsid proteins identified are herein described as SEQ ID NOS:1-7. The locations of the associated VP2 and VP3 regions are also herein described. The present invention is directed to (i) isolated AAV capsid proteins having at least 95%, 96%, 97%, 98% or 99% sequence identity to any of the VP1 capsid sequences of SEQ ID NOS:1-7, or the VP2 or VP3 regions of any of the capsid sequences of SEQ ID NOS:1-7, or (ii) isolated AAV capsid proteins comprising or consisting of any of the VP1 capsid sequences of SEQ ID NOS:1-7, or the VP2 or VP3 regions of any of the capsid sequences of SEQ ID NOS:1-7. The invention is also directed to an AAV particle that comprises any of the above described AAV capsid proteins, wherein the AAV particle further comprises either (i) a nucleic acid having AAV inverted terminal repeats and a transgene comprising a heterologous gene operably linked to regulatory sequences that direct expression of the heterologous gene in a host cell, or (ii) a nucleic acid comprising a heterologous gene operably linked to regulatory sequences that control expression of the heterologous gene in a host cell.

[00136] Neutralization of the novel AAV particles of the present invention by antibodies in human serum was also investigated. HEK293T cells were seeded in density $5x10^4$ cells/well and incubated overnight. Purified rAAVs were diluted to final titer of $2x10^6$ vg/uL and mixed with serial dilutions (0 – 10mg/mL) of IVIG for 1 hour. Recombinant AAVs were added onto

HEK293T cells using MOI of 1000 and incubated in 37°C. Seventy-two hours post-infection, IVIG neutralization was analyzed based on relative luciferase unit (RLU) reading. No etoposide was used in this study. The results are provided in Table 3. As expected, AAV2 transduction (positive control) was abolished by the addition of human IVIG (not shown). In contrast, certain of the novel AAVs tested exhibited IVIG resistant properties.

Table 3: IVIG Neutralization Data:

IVIG (mg/ml)	10	5	2.5	1.25	0.625	0.312	0.156	0.0781	0.039
Bba.45	268.0035	2460.32	25398.7	62037.8	120120.5	128270	163621.5	156841	174272
Bba.46	143.001	1030.049	6323.76	36486.35	92936.55	102423.15	179724	124453	156407.5
Bba.47	471.01865	1366.085	9060.61	42023.6	90104.35	123976.5	169643.5	152828.5	186058.5
Bba.49	271.004	1100.063	15301.9	65175.15	148917	159579.5	222658.5	337455.5	309530.5
Bba.50	188.0021	680.027	1574.105	6904.1	14449.3	19913.45	21366.1	26117.05	28455.65
Bba.51	1223.0695	1508.1235	10817.1	55625	103595.5	151814	153542	178776	255409

[00137] In addition, the IVIG neutralization assay was repeated to compare neutralization of the recombinant AAV particles having the capsids listed in Table 3 (denoted as "novel capsids", with the neutralization of recombinant AAV particles having the control capsids: AAV5, AAV8 and AAV9. As shown in in Fig. 2, the AAV having the capsids disclosed herein exhibited IVIG resistant properties.

[00138] In order to determine the tissue specific infectivity of the AAV capsids disclosed herein, AAV comprising each of the capsids and expressing the luciferase transgene were generated (AAV-RSV-egfp-T2A-Fluc2). Male Balb/C mice were purchased from Charles River Breeding Laboratories. A dose of 2 x 10^{13} vg/kg of AAV-RSV-egfp-T2A-Fluc2 vector was injected into the tail vein of 8 week old mice. At 3 and 5 weeks post injection, *in vivo* bioluminescent imaging was performed using an *in vivo* imagining device (IVIS Lumina LT obtained from PerkinElmer Inc., Waltham, MA). In brief, the mice were anesthetized with 2% isofluorane and oxygen. 150 µl of 30 mg/ml of RediJect D-Luciferin Bioluminescent Substrate was injected intraperitoneally. Ten minutes after substrate injection, the animals were imaged with the *in vivo* imaging device using its cooled charge-coupled device (CCD) camera. Images were takes in the dorsal positions of the animals. Anesthesia was maintained throughout the entire imaging session by isofluorane-oxygen delivery in the light-tight imaging chamber.

[00139] The mice were sacrificed after the imaging sessions at 5 weeks post AAV injection. Various organs were harvested and imaged using the imaging device. The measurement conditions were the same as those used for *in vivo* imaging. For imaging, a gray-scale photograph of the animals was acquired, followed by bioluminescence image acquisition. Image data was processed and analyzed using living image software version 4.5.2 (PerkinsElmer Waltham, MA). Regions of interest (ROIs) were traced surrounding each animal as well as individual organs to quantify the total flux (TF) (photons/second) being released by luciferase activity. Total flux activity is a proxy for AAV infectivity of each organ system and is shown for each AAV capsid in Figure 1. The data demonstrate that the novel capsids produce recombinant AAV that have a high degree of specificity for liver cells.

[00140] Additional experiments were carried out as described above using AAV having the following capsids disclosed herein (Bba.45, Bba.46, Bba.47, Bba.49, Bba.50 and Bba.51), and the transduction was compared with that AAV having the following control capsids: AAV5, AAV8 and AAV9. As shown in Fig. 3, the AAV having the novel capsids exhibit an increase in transduction efficiency compared to AAV having the capsids from AAV5 or AAV12. For example, the novel capsids have a 10-40% fold increase in transduction efficiency compared to AAV having the AAV5 capsid. Furthermore, Fig. 4 demonstrates that the AAV having the novel capsids have a significantly higher degree of specificity for liver cells compared to AAV5. In addition, the AAV having the novel capsids exhibited transduction of liver cells similar to that observed with AAV having AAV8 or AAV9 capsids.

Example 2

Evaluation of Evaluate Bio-distribution and Activity of Novel Naturally-Occurring Capsid Proteins

[00141] To evaluate bio-distribution and activity of the AAV capsids disclosed herein, AAV comprising the capsids, the choriogonadotropin subunit beta (cyno-CG-Beta) transgene under the ApoE-hAAT promoter were generated (AAV-ApoE-hAAT-Cyno-CG-Beta). Male C57BL/6J mice were purchased from Jackson Laboratories . A dose of 2 x 10^{13} vg/kg of AAV-ApoE-hAAT-Cyno-CG-Beta was injected into the tail vein of 8 week old mice. This study was carried out with AAV having the following capsid proteins: Bba-45, Bba-46, Bba-47, Bba-49, Bba-50 (denoted collectively as "novel capsids") and AAV5, AAV8, AAV-Rh10 and AAV-anc80L65

(denoted collectively as "control capsids"). The vehicle control was administration without an AAV vector.

[00142] At 5 weeks post injection, expression of the cyno-CG transgene was evaluated by measuring the plasma level of the bCG protein using mass spectrometry. As shown in Figure 5, the AAV having the novel capsids expressed the transgene at a level that is similar to the expression in AAV having the AAV5 capsid. However, the expression of bCG protein in the liver was increased in mice injected with AAV having the novel capsids compared to mice injected with AAV having the AAV5 capsid (Fig. 6). The number of DNA and RNA copies of the cyno-CG transgene in the liver of mice injected with AAV having the novel capsids were not significantly different than mice injected AAV having the AAV5 capsid (Fig. 7A-7C). The DNA and RNA data correlated well to the bCG protein data (see Fig. 8A-8B).

[00143] Novel capsids Bba-45 to Bba-50 did not lead to significantly higher transduction or transcript levels compared to the control capsids (AAV5, AAV8, AAV-Rh10 and AAV-anc80L65). However, when comparing all of the novel capsids, Bba-49 achieved the highest transcript and transduction levels. The Bba-49 capsid achieved about a 2-fold higher transcript levels (RNA) compared to the AAV5 capsid, but this difference is not significant (Fig.7B). The ratio of Bba-49 to AAV5 are set out in Table 4:

Table 4

Ratio between Bba49 and AAV5	
Protein	2.23
RNA	1.98
DNA	0.90

[00144] The expression of transgene bCG in hepatocytes was evaluated using immunohistochemistry. All of the novel capsids resulted in a higher percent of hepatocytes expressing bCG than AAV5 capsid. In addition, the transduction of hepatocytes by the AAV having the novel capsids were similar to the control capsids, AAV8, AAV-Rh10 and AAV-anc80L65 (Fig. 9 and 10).

[00145] AAV comprising the Bba-49 capsid, ApoE-hAAT promoter, the alanine glyoxalate amino transferase (AGXT) transgene were generated. A dose of 1E14 vg/kg of AAV-AGXT was injected into the tail vein of 3 to 4 week old $^{-}$AGXT/$^{-/-}$C57BL/6J male mice. The transduction of the hepatocytes was compared to the AAV having the AAV5 capsid from various sources. The expression of transgene AGXT in hepatocytes was evaluated using immunohistochemistry. As shown in Figure 11, the Bba-49 capsid resulted in a higher percent of hepatocytes expressing AGXT the AAV5 capsid tested. The AAV.Bba-49. AGXT transduced about 96% of the hepatocytes.