[19]	INTELLECTUAL PROPERTY PHILIPPINES			
[12]	INVENTION PUBLICATION			
[11]	Publication Number:	12014501487	Document Code:	B1
[22]	Publication Date:	7/12/2015		
[21]	Application Number:	12014501487	Document Code:	A
[22]	Date Filed:	26/6/2014		
[54]	Title:	A METHOD OF DETECTING AND/OR IDENTIFYING ADENO-ASSOCIATED VIRUS (AAV) SEQUENCES AND ISOLATING NOVEL SEQUENCES IDENTIFIED THEREBY		
[71]	Applicant(s):	UNIV PENNSYLVANIA		
[72]	Inventor(s):	GAO GUANGPING WILSON JAMES M ALVIRA MAURICIO		
[30]	Priority Data:	1/5/2002 US20020377066P		
[51]	International Class 8:	C12Q 20060101S I20051110RMEP; A61K 35/76 20060101ALI20180118DMJP; A61K 38/00 20060101ALI20051220RMJP; A61K 38/43 20060101ALI20051220RMJP; A61K 48/00 20060101ALI20051220RMJP; A61P 1/04 20060101ALI20051220RMJP; A61P 3/10 20060101ALI20051220RMJP; A61P 5/14 20060101ALI20051220RMJP; A61P 7/04 20060101ALI20051220RMJP; A61P 9/00 20060101ALI20051220RMJP; A61P 17/06 20060101ALI20051220RMJP; A61P 19/02 20060101ALI20051220RMJP; A61P 21/00 20060101ALI20051220RMJP; A61P 25/00 20060101ALI20051220RMJP; A61P 29/00 20060101ALI20051220RMJP; A61P 37/02 20060101ALI20051220RMJP; A61P 37/06 20060101ALI20051220RMJP; C07K 14/015 20060101A I20051008RMEP; C07K 14/075 20060101ALI20051220RMJP; C12N 1/15 20060101ALI20051220RMJP; C12N 1/19 20060101ALI20051220RMJP; C12N 1/21 20060101ALI20051220RMJP; C12N 5/10 20060101ALI20051220RMJP; C12N 7/00 20060101A I20051008RMEP; C12N 7/02 20060101A I20051008RMEP; C12N 7/04 20060101A I20051008RMEP; C12N 15/09 20060101AFI20051220RMJP; C12N 15/864 20060101A I20051008RMEP; C12Q 1/68 20060101ALI20051220RMJP; C12Q 1/70 20060101A I20051008RMEP; C12R 1/93 20060101ALN20051220RMJP; C07K 14/015 20060101AFI20180528BHPH; C12N 15/864 20060101ALI20180528BHPH; C12N 7/00 20060101ALI20180528BHPH; C12N 7/02 20060101ALI20180528BHPH; C12N 7/04 20060101ALI20180528BHPH;		
[57]	Abstract:	A method for detecting and isolating AAV sequences in a sample of DNA obtained from tissue or cells is provided. The invention further provides AAV sequences identified by this method, and vectors constructed using these sequences.		

another amino acid, there is amino acid sequence identity in at least about 95% to 99% of the aligned sequences. Preferably, the homology is over full-length sequence, 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.

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.

The term "percent sequence identity" or "identical" in the context of nucleic 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 genome, 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.

The AAV sequences 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 AAV sequences of the invention.

As described herein, the vectors of the invention containing 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 readministration and repeat gene therapy.

These and other embodiments and advantages of the invention are described in more detail below. As used throughout this specification and the claims, the terms "comprising" and "including" and their variants are inclusive of other components, elements, integers, steps and the like. Conversely, the term "consisting" and its variants is exclusive of other components, elements, integers, steps and the like.

I. Methods of the Invention

A. Detection of Sequences Via Molecular Cloning

In one aspect, the invention provides a method of detecting and/or identifying target nucleic acid sequences in a sample. This method is particularly well suited for detection of viral sequences which are

integrated into the chromosome of a cell, e.g., adeno-associated viruses (AAV) and retroviruses, among others. The specification makes reference to AAV, which is exemplified herein. However, based on this information, one of skill in the art may readily perform the methods of the invention on retroviruses [e.g., feline leukemia virus (FeLV), HTLVI and HTLVII], and lentivirinae [e.g., human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), equine infectious anemia virus, and spumavirinal)], among others. Further, the method of the invention may also be used for detection of other viral and non-viral sequences, whether integrated or non-integrated into the genome of the host cell.

As used herein, a sample is any source containing nucleic acids, e.g., tissue, tissue culture, cells, cell culture, and biological fluids including, without limitation, urine and blood. These nucleic acid sequences may be DNA or RNA from plasmids, natural DNA or RNA from any source, including bacteria, yeast, viruses, and higher organisms such as plants or animals. DNA or RNA is extracted from the sample by a variety of techniques known to those of skill in the art, such as those described by Sambrook, Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory). The origin of the sample and the method by which the nucleic acids are obtained for application of the method of the invention is not a limitation of the present invention. Optionally, the method of the invention can be performed directly on

the source of DNA, or on nucleic acids obtained (e.g., extracted) from a source.

The method of the invention involves subjecting a sample containing DNA to amplification via polymerase chain reaction (PCR) using a first set of primers specific for a first region of double-stranded nucleic acid sequences, thereby obtaining amplified sequences.

As used herein, each of the "regions" is predetermined based upon the alignment of the nucleic acid sequences of at least two serotypes (e.g., AAV) or strains (e.g., lentiviruses), and wherein each of said regions is composed of sequences having a 5' end which is highly conserved, a middle which is preferably, but necessarily, variable, and a 3' end which is highly conserved, each of these being conserved or variable relative to the sequences of the at least two aligned AAV serotypes. Preferably, the 5' and/or 3' end is highly conserved over at least about 9, and more preferably, at least 18 base pairs (bp). However, one or both of the sequences at the 5′ or 3′ end may be conserved over more than 18 bp, more than 25 bp, more than 30 bp, or more than 50 bp at the 5' end. With respect to the variable region, there is no requirement for conserved sequences, these sequences may be relatively conserved, or may have less than 90, 80, or 70% identity among the aligned serotypes or strains.

Each of the regions may span about 100 bp to about 10 kilobase pairs in length. However, it is particularly desirable that one of the regions is a "signature region", i.e., a region which is sufficiently unique to positively identify the

amplified sequence as being from the target source. For example, in one embodiment, the first region is about 250 bp in length, and is sufficiently unique among known AAV sequences, that it positively identifies the amplified region as being of AAV origin. Further, the variable sequences within this region are sufficiently unique that can be used to identify the serotype from which the amplified sequences originate. Once amplified (and thereby detected), the sequences can be identified by performing conventional restriction digestion and comparison to restriction digestion patterns for this region in any of AAV1, AAV2, AAV3, AAV4, AAV5, or AAV6, or that of AAV7, AAV10, AAV11, AAV12, or any of the other novel serotypes identified by the invention, which is predetermined and provided by the present invention.

Given the guidance provided herein, one of skill in the art can readily identify such regions among other integrated viruses to permit ready detection and identification of these sequences. Thereafter, an optimal set of generic primers located within the highly conserved ends can be designed and tested for efficient amplification of the selected region from samples. This aspect of the invention is readily adapted to a diagnostic kit for detecting the presence of the target sequence (e.g., AAV) and for identifying the AAV serotype, using standards which include the restriction patterns for the AAV serotypes described herein or isolated using the techniques described herein. For example, quick identification or molecular serotyping of PCR products can be accomplished by

digesting the PCR products and comparing restriction patterns.

Thus, in one embodiment, the "signature region" for AAV spans about bp 2800 to about 3200 of AAV 1 [SEQ ID NO:6], and corresponding base pairs in AAV 2, AAV3, AAV4, AAV5, and AAV6. More desirably, the region is about 250 bp, located within bp 2886 to about 3143 bp of AAV 1 [SEQ ID NO:6], and corresponding base pairs in AAV 2 [SEQ ID NO:7], AAV3 [SEQ ID NO8], and other AAV serotypes. See, Fig. 1. To permit rapid detection of AAV in the sample, primers which specifically amplify this signature region are utilized. However, the present invention is not limited to the exact sequences identified herein for the AAV signature region, as one of skill in the art may readily alter this region to encompass a shorter fragment, or a larger fragment of this signature region.

The PCR primers are generated using techniques known to those of skill in the art. Each of the PCR primer sets is composed of a 5′ primer and a 3′ primer. See, e.g., Sambrook et al, cited herein. The term "primer" refers to an oligonucleotide which acts as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced. The primer is preferably single stranded. However, if a double stranded primer is utilized, it is treated to separate its strands before being used to prepare extension products. The primers may be about 15 to 25 or more nucleotides, and preferably at least 18 nucleotides. However, for certain applications

shorter nucleotides, e.g., 7 to 15 nucleotides are utilized.

The primers are selected to be sufficiently complementary to the different strands of each specific sequence to be amplified to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the region being amplified. For example, a non-complementary nucleotide fragment may be attached to the 5′ end of the primer, with the remainder of the primer sequence being completely complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and form a template for synthesis of the extension product of the other primer.

The PCR primers for the signature region according to the invention are based upon the highly conserved sequences of two or more aligned sequences (e.g., two or more AAV serotypes). □The primers can accommodate less than exact identity among the two or more aligned AAV serotypes at the 5′ end or in the middle. However, the sequences at the 3′ end of the primers correspond to a region of two or more aligned AAV serotypes in which there is exact identity over at least five, preferably, over at least nine base pairs, and more preferably, over at least 18 base pairs at the 3′ end of the primers. Thus, the 3′ end of the primers is composed of sequences with 100% identity to the aligned sequences over at least five nucleotides. However, one

can optionally utilize one, two, or more degenerate nucleotides at the 3′ end of the primer.

For example, the primer set for the signature region of AAV was designed based upon a unique region within the AAV capsid, as follows. The 5′ primer was based upon nt 2867-2891 of AAV2 [SEQ ID NO:7], 5′-GGTAATTCCTCCGGAAATTGGCATT3′. See, Fig. 1. The 3′ primer was designed based upon nt 3096-3122 of AAV2 [SEQ ID NO:7], 5′- GACTCATCAACAACAACTGGGGATTC-3′. However, one of skill in the art may have readily designed the primer set based upon the corresponding regions of AAV 1, AAV3, AAV4, AAV5, AAV6, or based upon the information provided herein, AAV7, AAV10, AAV11, AAV12, or another novel AAV of the invention. In addition, still other primer sets can be readily designed to amplify this signature region, using techniques known to those of skill in the art.

B. Isolation of Target Sequences

As described herein, the present invention provides a first primer set which specifically amplifies the signature region of the target sequence, e.g., an AAV serotype, in order to permit detection of the target. In a situation in which further sequences are desired, e.g., if a novel AAV serotype is identified, the signature region may be extended. Thus, the invention may further utilize one or more additional primer sets.

Suitably, these primer sets are designed to include either the 5′ or 3′ primer of the first primer set and a second primer unique to the primer set, such

that the primer set amplifies a region 5′ or 3′ to the signature region which anneals to either the 5′ end or the 3′ end of the signature region. For example, a first primer set is composed of a 5′ primer, P1 and a 3′ primer P2 to amplify the signature region. In order to extend the signature region on its 3′ end, a second primer set is composed of primer P1 and a 3′ primer P4, which amplifies the signature region and contiguous sequences downstream of the signature region. In order to extend the signature region on its 5′ end, a third primer set is composed of a 5′ primer, P5, and primer P2, such that the signature region and contiguous sequences upstream of the signature region are amplified. These extension steps are repeated (or performed at the same time), as needed or desired. Thereafter, the products results from these amplification steps are fused using conventional steps to produce an isolated sequence of the desired length.

The second and third primer sets are designed, as with the primer set for the signature region, to amplify a region having highly conserved sequences among the aligned sequences. Reference herein to the term "second" or "third" primer set is for each of discussion only, and without regard to the order in which these primers are added to the reaction mixture, or used for amplification. The region amplified by the second primer set is selected so that upon amplification it anneals at its 5′ end to the 3′ end of the signature region. Similarly, the region amplified by the third primer set is selected so that upon amplification it anneals at its 3′ end anneals to the

5′ end of the signature region. Additional primer sets can be designed such that the regions which they amplify anneal to the either the 5′ end or the 3′ end of the extension products formed by the second or third primer sets, or by subsequent primer sets.

For example, where AAV is the target sequence, a first set of primers (P1 and P2) are used to amplify the signature region from the sample. In one desirable embodiment, this signature region is located within the AAV capsid. A second set of primers (P1 and P4) is used to extend the 3′ end of the signature region to a location in the AAV sequence which is just before the AAV 3′ ITR, i.e., providing an extension product containing the entire 3′ end of the AAV capsid when using the signature region as an anchor. In one embodiment, the P4 primer corresponds to nt 4435 to 4462 of AAV2 [SEQ ID NO:7], and corresponding sequences in the other AAV serotypes. This results in amplification of a region of about 1.6 kb, which contains the 0.25 kb signature region. A third set of primers (P3 and P2) is used to extend the 5′ end of signature region to a location in the AAV sequences which is in the 3′ end of the rep genes, i.e., providing an extension product containing the entire 5′ end of the AAV capsid when using the signature region as an anchor. In one embodiment, the P3 primer corresponds to nt 1384 to 1409 of AAV2 [SEQ ID NO:7], and corresponding sequences in the other AAV serotypes. This results in amplification of a region of about 1.7 kb, which contains the 0.25 kb signature region. Optionally, a fourth set of primers are used to further

extend the extension product containing the entire 5' end of the AAV capsid to also include the rep sequences. In one embodiment, the primer designated P5 corresponds to nt 108 to 133 of AAV2 [SEQ ID NO:7], and corresponding sequences in the other AAV serotypes and is used in conjunction with the P2 primer.

Following completion of the desired number of extension steps, the various extension products are fused, making use of the signature region as an anchor or marker, to construct an intact sequence. In the example provided herein, AAV sequences containing, at a minimum, an intact AAV cap gene are obtained. Larger sequences may be obtained, depending upon the number of extension steps performed.

Suitably, the extension products are assembled into an intact AAV sequence using methods known to those of skill in the art. For example, the extension products may be digested with DraIII, which cleaves at the DraIII site located within the signature region, to provide restriction fragments which are re-ligated to provide products containing (at a minimum) an intact AAV cap gene. However, other suitable techniques for assembling the extension products into an intact sequence may be utilized. See, generally, Sambrook et al, cited herein.

As an alternative to the multiple extension steps described above, another embodiment of the invention provides for direct amplification of a 3.1 kb fragment which allows isolation of full-length cap sequences. To directly amplify a 3.1 kb full-length cap fragment from NHP tissue and blood DNAs, two other highly

conserved regions were identified in AAV genomes for use in PCR amplification of large fragments. A primer within a conserved region located in the middle of the rep gene is utilized (AV1ns: 5' GCTGCGTCAACTGGACCAATGAGAAC 3', nt of SEQ ID NO:6) in combination with the 3' primer located in another conserved region downstream of the Cap gene (AV2cas: 5' CGCAGAGACCAAAGTTCAACTGAAACGA 3', SEQ ID NO: 7) for amplification of AAV sequences including the full-length AAV cap. Typically, following amplification, the products are cloned and sequence analysis is performed with an accuracy of ≥ 99.9%. Using this method, the inventors have isolated at least 50 capsid clones which have subsequently been characterized. Among them, 37 clones were derived from Rhesus macaque tissues (rh.1 - rh.37), 6 clones from cynomologous macaques (cy.1 - cy.6), 2 clones from Baboons (bb.1 and bb.2) and 5 clones from Chimps (ch.1 - ch.5). These clones are identified elsewhere in the specification, together with the species of animal from which they were identified and the tissues in that animal these novel sequences have been located.

C. Alternative method for isolating novel AAV

In another aspect, the invention provides an alternative method for isolating novel AAV from a cell. This method involves infecting the cell with a vector which provides helper functions to the AAV; isolating infectious clones containing AAV; sequencing the isolated AAV; and comparing the sequences of the isolated AAV to known AAV serotypes, whereby

differences in the sequences of the isolated AAV and known AAV serotypes indicates the presence of a novel AAV.

In one embodiment, the vector providing helper functions provides essential adenovirus functions, including, e.g., E1a, E1b, E2a, E4ORF6. In one embodiment, the helper functions are provided by an adenovirus. The adenovirus may be a wild-type adenovirus, and may be of human or non-human origin, preferably non-human primate (NHP) origin. The DNA sequences of a number of adenovirus types are available from Genbank, including type Ad5 [Genbank Accession No. M73260]. The adenovirus sequences may be obtained from any known adenovirus serotype, such as serotypes 2, 3, 4, 7, 12 and 40, and further including any of the presently identified human types [see, e.g., Horwitz, cited above]. Similarly adenoviruses known to infect non-human animals (e.g., chimpanzees) may also be employed in the vector constructs of this invention. See, e.g., US Patent No. 6,083,716. In addition to wild-type adenoviruses, recombinant viruses or non-viral vectors (e.g., plasmids, episomes, etc.) carrying the necessary helper functions may be utilized. Such recombinant viruses are known in the art and may be prepared according to published techniques. See, e.g., US Patent No. 5,871,982 and US Patent 6,251,677, which describe a hybrid Ad/AAV virus. The selection of the adenovirus type is not anticipated to limit the following invention. A variety of adenovirus strains are available from the American Type Culture Collection, Manassas, Virginia, or available by request

from a variety of commercial and institutional sources. Further, the sequences of many such strains are available from a variety of databases including, e.g., PubMed and GenBank.

In another alternative, infectious AAV may be isolated using genome walking technology (Siebert *et al.*, 1995, *Nucleic Acid Research*, **23**:1087-1088, Friezner-Degen *et al.*, 1986, *J. Biol. Chem.* **261**:6972-6985, BD Biosciences Clontech, Palo Alto, CA). Genome walking is particularly well suited for identifying and isolating the sequences adjacent to the novel sequences identified according to the method of the invention. For example, this technique may be useful for isolating inverted terminal repeat (ITRs) of the novel AAV serotype, based upon the novel AAV capsid and/or rep sequences identified using the methods of the invention. This technique is also useful for isolating sequences adjacent to other AAV and non-AAV sequences identified and isolated according to the present invention. See, Examples 3 and 4.

The methods of the invention may be readily used for a variety of epidemiology studies, studies of biodistribution, monitoring of gene therapy via AAV vectors and vector derived from other integrated viruses. Thus, the methods are well suited for use in pre-packaged kits for use by clinicians, researchers, and epidemiologists.

II. Diagnostic Kit

In another aspect, the invention provides a diagnostic kit for detecting the presence of a known or

unknown adeno-associated virus (AAV) in a sample. Such a kit may contain a first set of 5' and 3' PCR primers specific for a signature region of the AAV nucleic acid sequence. Alternatively, or additionally, such a kit can contain a first set of 5′ and 3′ PCR primers specific for the 3.1 kb fragment which includes the full-length AAV capsid nucleic acid sequence identified herein (e.g., the AV1ns and AV2cas primers.) Optionally, a kit of the invention may further contain two or more additional sets of 5' and 3' primers, as described herein, and/or PCR probes. These primers and probes are used according to the present invention amplify signature regions of each AAV serotype, e.g., using quantitative PCR.

The invention further provides a kit useful for identifying an AAV serotype detected according to the method of the invention and/or for distinguishing novel AAV from known AAV. Such a kit may further include one or more restriction enzymes, standards for AAV serotypes providing their "signature restriction enzyme digestions analyses", and/or other means for determining the serotype of the AAV detected.

In addition, kits of the invention may include, instructions, a negative and/or positive control, containers, diluents and buffers for the sample, indicator charts for signature comparisons, disposable gloves, decontamination instructions, applicator sticks or containers, and sample preparator cups, as well as any desired reagents, including media, wash reagents and concentration reagents. Such reagents may be readily selected from among the reagents described

herein, and from among conventional concentration reagents. In one desirable embodiment, the wash reagent is an isotonic saline solution which has been buffered to physiologic pH, such as phosphate buffered saline (PBS); the elution reagent is PBS containing 0.4 M NaCl, and the concentration reagents and devices. For example, one of skill in the art will recognize that reagents such as polyethylene glycol (PEG), or NH_4SO_4 may be useful, or that devices such as filter devices. For example, a filter device with a 100 K membrane would concentrate rAAV.

The kits provided by the present invention are useful for performing the methods described herein, and for study of biodistribution, epidemiology, mode of transmission of novel AAV serotypes in human and NHPs.

Thus, the methods and kits of the invention permit detection, identification, and isolation of target viral sequences, particularly integrated viral sequences. The methods and kits are particularly well suited for use in detection, identification and isolation of AAV sequences, which may include novel AAV serotypes.

In one notable example, the method of the invention facilitated analysis of cloned AAV sequences by the inventors, which revealed heterogeneity of proviral sequences between cloned fragments from different animals, all of which were distinct from the known six AAV serotypes, with the majority of the variation localized to hypervariable regions of the capsid protein. Surprising divergence of AAV sequences was noted in clones isolated from single tissue

sources, such as lymph node, from an individual rhesus monkey. This heterogeneity is best explained by apparent evolution of AAV sequence within individual animals due, in part, to extensive homologous recombination between a limited number of co-infecting parenteral viruses. These studies suggest sequence evolution of widely disseminated virus during the course of a natural AAV infection that presumably leads to the formation of swarms of quasispecies which differ from one another in the array of capsid hypervariable regions. This is the first example of rapid molecular evolution of a DNA virus in a way that formerly was thought to be restricted to RNA viruses.

Sequences of several novel AAV serotypes identified by the method of the invention and characterization of these serotypes is provided.

III. Novel AAV Serotypes

A. Nucleic Acid Sequences

Nucleic acid sequences of novel AAV serotypes identified by the methods of the invention are provided. See, SEQ ID NO:1, 9 - 59, and 117 - 120, which are incorporated by reference herein. See also, Fig. 1 and the sequence listing.

For novel serotype AAV7, the full-length sequences, including the AAV 5' ITRs, capsid, rep, and AAV 3' ITRs are provided in SEQ ID NO:1.

For other novel AAV serotypes of the invention, the approximately 3.1 kb fragment isolated according to the method of the invention is provided. This fragment contains sequences encoding full-length

capsid protein and all or part of the sequences encoding the rep protein. These sequences include the clones identified below.

For still other novel AAV serotypes, the signature region encoding the capsid protein is provided. For example, the AAV10 nucleic acid sequences of the invention include those illustrated in Fig. 1 [See, SEQ ID NO:117, which spans 255 bases]. The AAV11 nucleic acid sequences of the invention include the DNA sequences illustrated in Fig. 1 [See, SEQ ID NO:118 which spans 258 bases]. The AAV12 nucleic acid sequences of the invention include the DNA sequences illustrated in Fig. 1 [See, SEQ ID NO:119, which consists of 255 bases]. Using the methodology described above, further AAV10, AAV11 and AAV12 sequences can be readily identified and used for a variety of purposes, including those described for AAV7 and the other novel serotypes herein.

Figure 1 provides the non-human primate (NHP) AAV nucleic acid sequences of the invention in an alignment with the previously published AAV serotypes, AAV 1 [SEQ ID NO:6], AAV2 [SEQ ID NO:7], and AAV3 [SEQ ID NO:8]. These novel NHP sequences include those provided in the following Table I, which are identified by clone number:

Table 1

AAV Cap Sequence	Clone Number	Source		
		Species	Tissue	SEQ ID NO (DNA)
Rh.1	Clone 9 (AAV9)	Rhesus	Heart	5
Rh.2	Clone 43.1	Rhesus	MLN	39
Rh.3	Clone 43.5	Rhesus	MLN	40
Rh.4	Clone 43.12	Rhesus	MLN	41
Rh.5	Clone 43.20	Rhesus	MLN	42
Rh.6	Clone 43.21	Rhesus	MLN	43
Rh.7	Clone 43.23	Rhesus	MLN	44

Table 1 (cont'd)

Rh.8	Clone 43.25	Rhesus	MLN	45
Rh.9	Clone 44.1	Rhesus	Liver	46
Rh.10	Clone 44.2	Rhesus	Liver	59
Rh.11	Clone 44.5	Rhesus	Liver	47
Rh.12	Clone 42.1B	Rhesus	MLN	30
Rh.13	42.2	Rhesus	MLN	9
Rh.14	Clone 42.3A	Rhesus	MLN	32
Rh.15	Clone 42.3B	Rhesus	MLN	36
Rh.16	Clone 42.4	Rhesus	MLN	33
Rh.17	Clone 42.5A	Rhesus	MLN	34
Rh.18	Clone 42.5B	Rhesus	MLN	29
Rh.19	Clone 42.6B	Rhesus	MLN	38
Rh.20	Clone 42.8	Rhesus	MLN	27
Rh.21	Clone 42.10	Rhesus	MLN	35

A METHOD OF DETECTING AND/OR IDENTIFYING ADENO-ASSOCIATED VIRUS (AAV) SEQUENCES AND ISOLATING NOVEL SEQUENCES IDENTIFIED THEREBY

This application is a division of Application NO. 1-2010-501546 filed on 07 July 2010.

BACKGROUND OF THE INVENTION

Adeno-associated virus (AAV), a member of the Parvovirus family, is a small nonenveloped, icosahedral virus with single-stranded linear DNA genomes of 4.7 kilobases (kb) to 6 kb. AAV is assigned to the genus, Dependovirus, because the virus was discovered as a contaminant in purified adenovirus stocks. AAV's life cycle includes a latent phase at which AAV genomes, after infection, are site specifically integrated into host chromosomes and an infectious phase in which, following either adenovirus or herpes simplex virus infection, the integrated genomes are subsequently rescued, replicated, and packaged into infectious viruses. The properties of non-pathogenicity, broad host range of infectivity, including non-dividing cells, and potential site-specific chromosomal integration make AAV an attractive tool for gene transfer.

Recent studies suggest that AAV vectors may be the preferred vehicle for gene therapy. To date, there have been 6 different serotypes of AAVs isolated from human or non-human primates (NHP) and well characterized. Among them, human serotype 2 is the first AAV that was developed as a gene transfer vector; it has been widely used for efficient gene transfer experiments in different target tissues and animal models. Clinical trials of the experimental

Rh.22	Clone 42.11	Rhesus	MLN	37
Rh.23	Clone 42.12	Rhesus	MLN	58
Rh.24	Clone 42.13	Rhesus	MLN	31
Rh.25	Clone 42.15	Rhesus	MLN	28
Rh.26	Clone 223.2	Rhesus	Liver	49
Rh.27	Clone 223.4	Rhesus	Liver	50
Rh.28	Clone 223.5	Rhesus	Liver	51
Rh.29	Clone 223.6	Rhesus	Liver	52
Rh.30	Clone 223.7	Rhesus	Liver	53
Rh.31	Clone 223.10	Rhesus	Liver	48
Rh.32	Clone C1	Rhesus	Spleen, Duo, Kid & Liver	19
Rh.33	Clone C3	Rhesus		20
Rh.34	Clone C5	Rhesus		21
Rh.35	Clone F1	Rhesus	Liver	22

Rh.36	Clone F3	Rhesus		23
Rh.37	Clone F5	Rhesus		24
Cy.1	Clone 1.3	Cyno	Blood	14
Cy.2	Clone 13.3B	Cyno	Blood	15
Cy.3	Clone 24.1	Cyno	Blood	16
Cy.4	Clone 27.3	Cyno	Blood	17
Cy.5	Clone 7.2	Cyno	Blood	18
Cy.6	Clone 16.3	Cyno	Blood	10

Table 1 (cont'd)

bb.1	Clone 29.3	Baboon	Blood	11
bb.2	Clone 29.5	Baboon	Blood	13
Ch.1	Clone A3.3	Chimp	Blood	57
Ch.2	Clone A3.4	Chimp	Blood	54
Ch.3	Clone A3.5	Chimp	Blood	55
Ch.4	Clone A3.7	Chimp	Blood	56

A novel NHP clone was made by splicing capsids fragments of two chimp adenoviruses into an AAV2 rep construct. This new clone, A3.1, is also termed Ch.5 [SEQ ID NO:20]. Additionally, the present invention includes two human AAV sequences, termed H6 [SEQ ID NO:25] and H2 [SEQ ID NO:26].

The AAV nucleic acid sequences of the invention further encompass the strand which is complementary to the strands provided in the sequences provided in Fig. 1 and the Sequence Listing [SEQ ID NO:1, 9 - 59, 117 - 120], nucleic acid sequences, as well as the RNA and cDNA sequences corresponding to the sequences provided in Fig. 1 and the Sequence Listing [SEQ ID NO:1, 9 - 59, 117-120], and their complementary strands. Also included in the nucleic acid sequences of the invention are natural variants and engineered modifications of the sequences of Fig1 and the Sequence Listing [SEQ ID NO:1, 9 - 59, 117-120], and their complementary strands. 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.

Further included in this invention are nucleic acid sequences which are greater than 85%, preferably at least about 90%, more preferably at least about 95%, and most preferably at least about 98 to 99% identical or homologous to the sequences of the invention, including Fig. 1 and the Sequence Listing [SEQ ID NO:1, 9 - 59, 117-120]. These terms are as defined herein.

Also included within the invention are fragments of the novel AAV sequences identified by the method described herein. Suitable fragments are at least 15 nucleotides in length, and encompass functional fragments, i.e., fragments which are of biological interest. In one embodiment, these fragments are fragments of the novel sequences of Fig. 1 and the Sequence Listing [SEQ ID NO:1, 9 - 59, 117-120], their complementary strands, cDNA and RNA complementary thereto.

Examples of suitable fragments are provided with respect to the location of these fragments on AAV1, AAV2, or AAV7. However, using the alignment provided herein (obtained using the Clustal W program at default settings), or similar techniques for generating an alignment with other novel serotypes of the invention, one of skill in the art can readily identify the precise nucleotide start and stop codons for desired fragments.

Examples of suitable fragments include the sequences encoding the three variable proteins (vp) of the AAV capsid which are alternative splice variants: vp1 [e.g., nt 825 to 3049 of AAV7, SEQ ID NO: 1]; vp2 [e.g., nt 1234 - 3049 of AAV7, SEQ ID NO: 1]; and vp 3 [e.g., nt 1434 - 3049 of AAV7, SEQ ID NO:1]. It is notable that AAV7 has an unusual GTG start codon. With the exception of a few house-keeping genes, such a start codon has not previously been reported in DNA viruses. The start codons for vp1, vp2 and vp3 for other AAV serotypes have been believed to be such that they permit the cellular mechanism of the host cell in

which they reside to produce vp1, vp2 and vp3 in a ratio of 10%:10%:80%, respectively, in order to permit efficient assembly of the virion. However, the AAV7 virion has been found to assemble efficiently even with this rare GTG start codon. Thus, the inventors anticipate this it is desirable to alter the start codon of the vp3 of other AAV serotypes to contain this rare GTG start codon, in order to improve packaging efficiency, to alter the virion structure and/or to alter location of epitopes (e.g., neutralizing antibody epitopes) of other AAV serotypes. The start codons may be altered using conventional techniques including, e.g., site directed mutagenesis. Thus, the present invention encompasses altered AAV virions of any selected serotype, composed of a vp 3, and/or optionally, vp 1 and/or vp2 having start codons altered to GTG.

Other suitable fragments of AAV, include a fragment containing the start codon for the AAV capsid protein [e.g., nt 468 to 3090 of AAV7, SEQ ID NO:1, nt 725 to 3090 of AAV7, SEQ ID NO: 1, and corresponding regions of the other AAV serotypes]. Still other fragments of AAV7 and the other novel AAV serotypes identified using the methods described herein include those encoding the rep proteins, including *rep* 78 [e.g., initiation codon 334 of Fig 1 for AAV7], *rep* 68 [initiation codon nt 334 of Fig. 1 for AAV7], *rep* 52 [initiation codon 1006 of Fig. 1 for AAV7], and *rep* 40 [initiation codon 1006 of Fig. 1 for AAV7] Other fragments of interest may include the AAV 5' inverted terminal repeats ITRs, [nt 1 to 107 of Fig. 1 for

AAV7]; the AAV 3' ITRs [nt 4704 to 4721 of Fig. 1 for AAV7], P19 sequences, AAV P40 sequences, the rep binding site, and the terminal resolute site (TRS). Still other suitable fragments will be readily apparent to those of skill in the art. The corresponding regions in the other novel serotypes of the invention can be readily determined by reference to Figure 1, or by utilizing conventional alignment techniques with the sequences provided herein.

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 serotypes of the invention. Thus, the invention includes nucleic acid sequences which encode the following novel AAV amino acid sequences: C1 [SEQ ID NO:60], C2 [SEQ ID NO:61], C5 [SEQ ID NO:62], A3-3 [SEQ ID NO:66], A3-7 [SEQ ID NO:67], A3-4 [SEQ ID NO:68], A3-5 [SEQ ID NO: 69], 3.3b [SEQ ID NO: 62], 223.4 [SEQ ID NO: 73], 223-5 [SEQ ID NO:74], 223-10 [SEQ ID NO:75], 223-2 [SEQ ID NO:76], 223-7 [SEQ ID NO: 77], 223-6 [SEQ ID NO: 78], 44-1 [SEQ ID NO: 79], 44-5 [SEQ ID NO:80], 44-2 [SEQ ID NO:81], 42-15 [SEQ ID NO: 84], 42-8 [SEQ ID NO: 85], 42-13 [SEQ ID NO:86], 42-3A [SEQ ID NO:87], 42-4 [SEQ ID NO:88], 42-5A [SEQ ID NO:89], 42-1B [SEQ ID NO:90], 42-5B [SEQ ID NO:91], 43-1 [SEQ ID NO: 92], 43-12 [SEQ ID NO: 93], 43-5 [SEQ ID NO:94], 43-21 [SEQ ID NO:96], 43-25 [SEQ ID NO: 97], 43-20 [SEQ ID NO:99], 24.1 [SEQ ID NO: 101], 42.2 [SEQ ID NO:102], 7.2 [SEQ ID NO: 103], 27.3 [SEQ ID NO: 104], 16.3 [SEQ

ID NO: 105], 42.10 [SEQ ID NO: 106], 42-3B [SEQ ID NO: 107], 42-11 [SEQ ID NO: 108], F1 [SEQ ID NO: 109], F5 [SEQ ID NO: 110], F3 [SEQ ID NO:111], 42-6B [SEQ ID NO: 112], and/or 42-12 [SEQ ID NO: 113], and artificial AAV serotypes generated using these sequences and/or unique fragments thereof.

As used herein, artificial AAV serotypes include, without limitation, AAV with a non-naturally occurring capsid protein. Such an artificial capsid may be generated by any suitable technique, using a novel AAV sequence of the invention (e.g., a fragment of a vp1 capsid protein) in combination with heterologous sequences which may be obtained from another AAV serotype (known or novel), non-contiguous portions of the same AAV serotype, from a non-AAV viral source, or from a non-viral source. An artificial AAV serotype may be, without limitation, a chimeric AAV capsid, a recombinant AAV capsid, or a "humanized" AAV capsid.

B. AAV Amino Acid Sequences, Proteins and Peptides

The invention provides proteins and fragments thereof which are encoded by the nucleic acid sequences of the novel AAV serotypes identified herein, including, e.g., AAV7 [nt 825 to 3049 of AAV7, SEQ ID NO: 1] the other novel serotypes provided herein. Thus, the capsid proteins of the novel serotypes of the invention, including: H6 [SEQ ID NO: 25], H2 [SEQ ID NO: 26], 42-2 [SEQ ID NO:9], 42-8 [SEQ ID NO:27], 42-15 [SEQ ID NO:28], 42-5b [SEQ ID NO: 29], 42-1b [SEQ ID NO:30]; 42-13 [SEQ ID NO: 31], 42-3a [SEQ ID NO: 32],

42-4 [SEQ ID NO:33], 42-5a [SEQ ID NO: 34], 42-10 [SEQ ID NO:35], 42-3b [SEQ ID NO: 36], 42-11 [SEQ ID NO: 37], 42-6b [SEQ ID NO:38], 43-1 [SEQ ID NO: 39], 43-5 [SEQ ID NO: 40], 43-12 [SEQ ID NO:41], 43-20 [SEQ ID NO:42], 43-21 [SEQ ID NO: 43], 43-23 [SEQ ID NO:44], 43-25 [SEQ ID NO: 45], 44.1 [SEQ ID NO:47], 44.5 [SEQ ID NO:47], 223.10 [SEQ ID NO:48], 223.2 [SEQ ID NO:49], 223.4 [SEQ ID NO:50], 223.5 [SEQ ID NO: 51], 223.6 [SEQ ID NO: 52], 223.7 [SEQ ID NO: 53], A3.4 [SEQ ID NO: 54], A3.5 [SEQ ID NO:55], A3.7 [SEQ ID NO: 56], A3.3 [SEQ ID NO:57], 42.12 [SEQ ID NO: 58], and 44.2 [SEQ ID NO: 59], can be readily generated using conventional techniques from the open reading frames provided for the above-listed clones.

The invention further encompasses AAV serotypes generated using sequences of the novel AAV serotypes of the invention, which are generated using synthetic, recombinant or other techniques known to those of skill in the art. The invention is not limited to novel AAV amino acid sequences, peptides and proteins expressed from the novel 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 C1 [SEQ ID NO:60], C2 [SEQ ID NO:61], C5 [SEQ ID NO:62], A3-3 [SEQ ID NO:66], A3-7 [SEQ ID NO:67], A3-4 [SEQ ID NO:68], A3-5 [SEQ ID NO: 69], 3.3b [SEQ ID NO: 62], 223.4 [SEQ ID NO: 73], 223-5 [SEQ ID NO:74], 223-10 [SEQ ID NO:75], 223-2 [SEQ ID NO:76], 223-7 [SEQ ID NO: 77], 223-6 [SEQ ID NO:

78], 44-1 [SEQ ID NO: 79], 44-5 [SEQ ID NO:80], 44-2 [SEQ ID NO:81], 42-15 [SEQ ID NO: 84], 42-8 [SEQ ID NO: 85], 42-13 [SEQ ID NO:86], 42-3A [SEQ ID NO:87], 42-4 [SEQ ID NO:88], 42-5A [SEQ ID NO:89], 42-1B [SEQ ID NO:90], 42-5B [SEQ ID NO:91], 43-1 [SEQ ID NO: 92], 43-12 [SEQ ID NO: 93], 43-5 [SEQ ID NO:94], 43-21 [SEQ ID NO:96], 43-25 [SEQ ID NO: 97], 43-20 [SEQ ID NO:99], 24.1 [SEQ ID NO: 101], 42.2 [SEQ ID NO:102], 7.2 [SEQ ID NO: 103], 27.3 [SEQ ID NO: 104], 16.3 [SEQ ID NO: 105], 42.10 [SEQ ID NO: 106], 42-3B [SEQ ID NO: 107], 42-11 [SEQ ID NO: 108], F1 [SEQ ID NO: 109], F5 [SEQ ID NO: 110], F3 [SEQ ID NO:111], 42-6B [SEQ ID NO: 112], and/or 42-12 [SEQ ID NO: 113] by be readily generated using a variety of techniques.

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, NY). 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.

Particularly desirable proteins include the AAV capsid proteins, which are encoded by the nucleotide sequences identified above. The sequences of many of the capsid proteins of the invention are provided in an alignment in Fig. 2 and/or in the Sequence Listing, SEQ

ID NO: 2 and 60 to 115, which is incorporated by reference herein. The AAV capsid is composed of three proteins, vp1, vp2 and vp3, which are alternative splice variants. The full-length sequence provided in these figures is that of vp1. Based on the numbering of the AAV7 capsid [SEQ ID NO:2], the sequences of vp2 span amino acid 138 - 737 of AAV7 and the sequences of vp3 span amino acids 203 - 737 of AAV7. With this information, one of skill in the art can readily determine the location of the vp2 and vp3 proteins for the other novel serotypes of the invention.

Other desirable proteins and fragments of the capsid protein include the constant and variable regions, located between hypervariable regions (HPV) and the sequences of the HPV regions themselves. An algorithm developed to determine areas of sequence divergence in AAV2 has yielded 12 hypervariable regions (HVR) of which 5 overlap or are part of the four previously described variable regions. [Chiorini *et al*, *J. Virol*, **73**:1309-19 (1999); Rutledge *et al*, *J. Virol.*, **72**:309-319] Using this algorithm and/or the alignment techniques described herein, the HVR of the novel AAV serotypes are determined. For example, with respect to the number of the AAV2 vp1 [SEQ ID NO:70], the HVR are located as follows: HVR1, aa 146-152; HVR2, aa 182-186; HVR3, aa 262-264; HVR4, aa 381-383; HVR5, aa 450-474; HVR6, aa 490-495; HVR7, aa500-504; HVR8, aa 514-522; HVR9, aa 534-555; HVR10, aa 581-594; HVR11, aa 658-667; and HVR12, aa 705-719. Utilizing an alignment prepared in accordance with conventional methods and the novel sequences provided herein [See,

application of AAV2 based vectors to some human disease models are in progress, and include such diseases as cystic fibrosis and hemophilia B.

What are desirable are AAV-based constructs for gene delivery.

SUMMARY OF THE INVENTION

In one aspect, the invention provides a novel method of detecting and identifying AAV sequences from cellular DNAs of various human and non-human primate (NHP) tissues using bioinformatics analysis, PCR based gene amplification and cloning technology, based on the nature of latency and integration of AAVs in the absence of helper virus co-infection.

In another aspect, the invention provides method of isolating novel AAV sequences detected using the above described method of the invention. The invention further comprises methods of generating vectors based upon these novel AAV serotypes, for serology and gene transfer studies solely based on availability of capsid gene sequences and structure of rep/cap gene junctions.

In still another aspect, the invention provides a novel method for performing studies of serology, epidemiology, biodistribution and mode of transmission, using reagents according to the invention, which include generic sets of primers/probes and quantitative real time PCR.

In yet another aspect, the invention provides a method of isolating complete and infectious genomes of novel AAV serotypes from cellular DNA of different origins using RACE and other molecular techniques.

e.g., Figure 2], one can readily determine the location of the HVR in the novel AAV serotypes of the invention. For example, utilizing Figure 2, one can readily determine that for AAV7 [SEQ ID NO:2]. HVR1 is located at aa 146 - 152; HVR2 is located at 182-187; HVR3 is located at aa 263-266, HVR4 is located at aa 383-385, HVR5 is located at aa 451-475; HVR6 is located at aa 491-496 of AAV7; HVR7 is located at aa 501-505; HVR8 is located at aa 513-521; HVR9 is located at 533-554; HVR10 is located at aa 583-596; HVR11 is located at aa 660-669; HVR12 is located at aa 707-721. Using the information provided herein, the HVRs for the other novel serotypes of the invention can be readily determined.

In addition, within the capsid, amino acid cassettes of identity have been identified. These cassettes are of particular interest, as they are useful in constructing artificial serotypes, e.g., by replacing a HVR1 cassette of a selected serotype with an HVR1 cassette of another serotype. Certain of these cassettes of identity are noted in Fig. 2. See, Fig. 2, providing the Clustal X alignment, which has a ruler is displayed below the sequences, starting at 1 for the first residue position. The line above the ruler is used to mark strongly conserved positions. Three characters (*, : , .) are used. "*" indicates positions which have a single, fully conserved residue. ":" indicates that a "strong" group is fully conserved "." Indicates that a "weaker" group is fully conserved. These are all the positively scoring groups that occur in the Gonnet Pam250 matrix. The strong groups are

defined as a strong score >0.5 and the weak groups are defined as weak score <0.5.

Additionally, examples of other suitable fragments of AAV capsids include, with respect to the numbering of AAV2 [SEQ ID NO:70], aa 24 - 42, aa 25 - 28; aa 81 - 85; aa133-165; aa 134 - 165; aa 137-143; aa 154-156; aa 194-208; aa 261-274; aa 262-274; aa 171-173; aa 413-417; aa 449-478; aa 494-525; aa 534-571; aa 581-601; aa 660-671; aa 709-723. Still other desirable fragments include, for example, in AAV7, amino acids 1 to 184 of SEQ ID NO:2, amino acids 199 to 259; amino acids 274 to 446; amino acids 603 to 659; amino acids 670 to 706; amino acids 724 to 736; aa 185 to 198; aa 260 to 273; aa447 to 477; aa495 to 602; aa660 to 669; and aa707 to 723. Still other desirable regions, based on the numbering of AAV7 [SEQ ID NO:2], are selected from among the group consisting of aa 185 to 198; aa 260 to 273; aa447 to 477;aa495 to 602; aa660 to 669; and aa707 to 723. 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.

Other desirable proteins are the AAV rep proteins [aa 1 to 623 of SEQ ID NO:3 for AAV7] and functional fragments thereof, including, e.g., aa 1 to 171, aa 172 to 372, aa 373 to 444, aa 445 to 623 of SEQ ID NO:3, among others. Suitably, such fragments are at least 8 amino acids in length. See, Fig. 3. Comparable regions can be identified in the proteins of the other

novel AAV of the invention, using the techniques described herein and those which are known in the art. In addition, fragments of other desired lengths may be readily utilized. Such fragments may be produced recombinantly or by other suitable means, e.g., chemical synthesis.

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.

IV. Production of rAAV with novel AAV capsids

The invention encompasses novel, wild-type AAV serotypes identified by the invention, the sequences of which wild-type AAV serotypes are free of DNA and/or cellular material with 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.

The molecules of the invention which contain sequences of a novel AAV serotype of the invention 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, NY.

In one embodiment, the vectors of the invention contain sequences encoding a novel AAV capsid of the invention (e.g., AAV7 capsid, AAV 44-2 (rh.10), an AAV10 capsid, an AAV11 capsid, an AAV12 capsid), or a fragment of one or more of these AAV capsids. Alternatively, the vectors may contain the capsid protein, or a fragment thereof, itself.

Optionally, vectors of the invention may contain sequences encoding AAV rep proteins. Such *rep* sequences may be from the same AAV serotype which is providing the cap sequences. 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 expressed from the same source as the *cap* sequences. In this embodiment, the *rep* sequences may be fused in frame to *cap* sequences of a different

AAV serotype to form a chimeric AAV vector. Optionally, the vectors of the invention further contain a minigene comprising a selected transgene which is flanked by AAV 5' ITR and AAV 3' ITR.

Thus, in one embodiment, the vectors described herein contain nucleic acid sequences encoding an intact AAV capsid which may be from a single AAV serotype (e.g., AAV7 or another novel AAV). Alternatively, these vectors contain sequences encoding artificial capsids which contain one or more fragments of the AAV7 (or another novel AAV) capsid fused to heterologous AAV or non-AAV capsid proteins (or fragments thereof). These artificial capsid proteins are selected from non-contiguous portions of the AAV7 (or another novel AAV) capsid or from capsids of other AAV serotypes. For example, it may be desirable to modify the coding regions of one or more of the AAV vp1, e.g., in one or more of the hypervariable regions (i.e., HPV1-12), or vp2, and/or vp3. In another example, it may be desirable to alter the start codon of the vp3 protein to GTG. 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).

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.

In one aspect, the invention provides a method of generating a recombinant adeno-associated virus (AAV) having an AAV serotype 7 (or another novel AAV) capsid, or a portion thereof. Such a method involves culturing a host cell which contains a nucleic acid sequence encoding an adeno-associated virus (AAV) serotype 7 (or another novel AAV) capsid protein, or fragment thereof, as defined herein; a functional rep gene; a minigene composed of, at a minimum, AAV inverted terminal repeats (ITRs) and a transgene; and sufficient helper functions to permit packaging of the minigene into the AAV7 (or another novel AAV) capsid protein.

The components required to be cultured in the host cell to package an AAV minigene in an AAV capsid may be provided to the host cell in *trans*. Alternatively, any one or more of the required components (e.g., minigene, *rep* sequences, *cap* sequences, and/or helper functions) may be provided by a stable host cell which has been engineered to contain one or more of the required components using methods known to those of skill in the art. Most suitably, such a stable host cell will contain the required component(s) under the control of an inducible promoter. However, the required component(s) may be under the control of a constitutive promoter. Examples of suitable inducible and constitutive promoters are provided herein, in the discussion of regulatory elements suitable for use with the transgene. In still another alternative, a selected stable host cell may contain selected

component(s) under the control of a constitutive promoter and other selected component(s) under the control of one or more inducible promoters. For example, a stable host cell may be generated which is derived from 293 cells (which contain E1 helper functions under the control of a constitutive promoter), but which contains the rep and/or cap proteins under the control of inducible promoters. Still other stable host cells may be generated by one of skill in the art.

The minigene, *rep* sequences, *cap* sequences, and helper functions required for producing the rAAV of the invention may be delivered to the packaging host cell in the form of any genetic element which transfer the sequences carried thereon. The selected genetic element may be delivered by any suitable method, including those described herein. 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, NY. Similarly, methods of generating rAAV virions are well known and the selection of a suitable method is not a limitation on the present invention. See, e.g., K. Fisher et al, J. Virol., 70:520-532 (1993) and US Patent 5,478,745.

A. The Minigene

The minigene is composed of, at a minimum, a transgene and its regulatory sequences, and 5′ and 3′ AAV inverted terminal repeats (ITRs). It is this minigene which is packaged into a capsid protein and delivered to a selected host cell.

1. The transgene

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.

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.

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.

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, or catalytic RNAs. Desirable RNA molecules include tRNA, dsRNA, ribosomal RNA, catalytic RNAs, and antisense RNAs. One example of a useful RNA sequence is a sequence which extinguishes expression of a targeted nucleic acid sequence in the treated animal.

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., M.L. Donnelly, *et al*, *J. Gen. Virol.*, **78**(Pt 1):13-21 (Jan 1997); Furler, S., *et al*, *Gene Ther.*, **8**(11):864-873 (June 2001); Klump H., *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. However, the selected transgene may encode any

In a further aspect, the invention provides a method of rescuing novel serotypes of AAV genomes from human and NHP cell lines using adenovirus helpers of different origins.

In still a further aspect, the invention provides novel AAV serotypes, vectors containing same, and methods of using same.

These and other aspects of the invention will be readily apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A through 1AAAR provide an alignment of the nucleic acid sequences encoding at least the cap proteins for the AAV serotypes. The full-length sequences including the ITRs, the rep region, and the capsid region are provided for novel AAV serotype 7 [SEQ ID NO:1], and for previously published AAV1 [SEQ IN NO:6], AAV2 [SEQ ID NO:7]; and AAV3 [SEQ ID NO:8]. Novel AAV serotypes AAV8 [SEQ ID NO:4] and AAV9 [SEQ ID NO:5] are the subject of co-filed applications. The other novel clones of the invention provided in this alignment include: 42-2 [SEQ ID NO:9], 42-8 [SEQ ID NO:27], 42-15 [SEQ ID NO:28], 42-5b [SEQ ID NO: 29], 42-1b [SEQ ID NO:30]; 42-13 [SEQ ID NO: 31], 42-3a [SEQ ID NO: 32], 42-4 [SEQ ID NO:33], 42-5a [SEQ ID NO: 34], 42-10 [SEQ ID NO:35], 42-3b [SEQ ID NO: 36], 42-11 [SEQ ID NO: 37], 42-6b [SEQ ID NO:38], 43-1 [SEQ ID NO: 39], 43-5 [SEQ ID NO: 40], 43-12 [SEQ ID NO:41], 43-20 [SEQ ID NO:42], 43-21 [SEQ ID NO: 43], 43-23 [SEQ ID NO:44], 43-25 [SEQ ID NO: 45], 44.1 [SEQ ID NO:47],

biologically active product or other product, e.g., a product desirable for study.

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.

2. Regulatory Elements

In addition to the major elements identified above for the minigene, the vector also includes conventional control elements necessary 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.

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.

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 EF1α 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 promoters 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)]. Still 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.

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.

Another embodiment of the transgene includes a transgene 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 *et al.*, *J. Immunol.*, **161**:1063-8 (1998); immunoglobulin heavy chain; T cell receptor α 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.

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].

The combination of the transgene, promoter/enhancer, and 5′ and 3′ ITRs is referred to as a "minigene" for ease of reference herein. Provided with the teachings of this invention, the design of such a minigene can be made by resort to conventional techniques.

3. Delivery of the Minigene to a Packaging Host Cell

The minigene can be carried on any suitable vector, e.g., a plasmid, which is delivered to a host cell. The plasmids useful in this invention may be engineered such that they are suitable for replication and, optionally, integration in prokaryotic cells, mammalian cells, or both. These plasmids (or other vectors carrying the 5' AAV ITR-heterologous molecule-3'ITR) contain sequences permitting replication of the minigene in eukaryotes and/or prokaryotes and selection markers for these systems. Selectable markers or reporter genes may include sequences encoding geneticin, hygromicin or purimycin resistance, among others. The plasmids may also contain certain selectable reporters or marker genes that can be used to signal the presence of the vector in bacterial cells, such as ampicillin resistance. Other components of the plasmid may include an origin of replication and an amplicon, such as the amplicon system employing the Epstein Barr virus nuclear antigen. This amplicon system, or other similar amplicon components permit high copy episomal replication in the cells. Preferably, the molecule carrying the minigene is

transfected into the cell, where it may exist transiently. Alternatively, the minigene (carrying the 5' AAV ITR-heterologous molecule-3' ITR) may be stably integrated into the genome of the host cell, either chromosomally or as an episome. In certain embodiments, the minigene may be present in multiple copies, optionally in head-to-head, head-to-tail, or tail-to-tail concatamers. Suitable transfection techniques are known and may readily be utilized to deliver the minigene to the host cell.

Generally, when delivering the vector comprising the minigene by transfection, the vector is delivered in an amount from about 5 μg to about 100 μg DNA, and preferably about 10 to about 50 μg DNA to about 1 x 10^4 cells to about 1 x 10^{13} cells, and preferably about 10^5 cells. However, the relative amounts of vector DNA to host cells may be adjusted, taking into consideration such factors as the selected vector, the delivery method and the host cells selected.

B. *Rep* and *Cap* Sequences

In addition to the minigene, the host cell contains the sequences which drive expression of the novel AAV capsid protein (e.g., AAV7 or other novel AAV capsid or an artificial capsid protein comprising a fragment of one or more of these capsids) in the host cell and rep sequences of the same serotype as the serotype of the AAV ITRs found in the minigene. The AAV *cap* and *rep* sequences may be independently obtained from an AAV source as described above and may be introduced into the host cell in any manner known to

one in the art as described above. Additionally, when pseudotyping a novel AAV capsid of the invention, the sequences encoding each of the essential rep proteins may be supplied by the same AAV serotype, or the sequences encoding the rep proteins may be supplied by different AAV serotypes (e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, or one of the novel serotypes identified herein). For example, the *rep*78/68 sequences may be from AAV2, whereas the *rep*52/40 sequences may from AAV1.

In one embodiment, the host cell stably contains the capsid protein under the control of a suitable promoter, such as those described above. Most desirably, in this embodiment, the capsid protein is expressed under the control of an inducible promoter. In another embodiment, the capsid protein is supplied to the host cell in *trans*. When delivered to the host cell in *trans*, the capsid protein may be delivered via a plasmid which contains the sequences necessary to direct expression of the selected capsid protein in the host cell. Most desirably, when delivered to the host cell in *trans*, the plasmid carrying the capsid protein also carries other sequences required for packaging the rAAV, e.g., the *rep* sequences.

In another embodiment, the host cell stably contains the *rep* sequences under the control of a suitable promoter, such as those described above. Most desirably, in this embodiment, the essential rep proteins are expressed under the control of an inducible promoter. In another embodiment, the rep proteins are supplied to the host cell in *trans*. When

delivered to the host cell in *trans*, the rep proteins may be delivered via a plasmid which contains the sequences necessary to direct expression of the selected rep proteins in the host cell. Most desirably, when delivered to the host cell in *trans*, the plasmid carrying the capsid protein also carries other sequences required for packaging the rAAV, e.g., the *rep* and *cap* sequences.

Thus, in one embodiment, the *rep* and *cap* sequences may be transfected into the host cell on a single nucleic acid molecule and exist stably in the cell as an episome. In another embodiment, the *rep* and *cap* sequences are stably integrated into the genome of the cell. Another embodiment has the *rep* and *cap* sequences transiently expressed in the host cell. For example, a useful nucleic acid molecule for such transfection comprises, from 5' to 3', a promoter, an optional spacer interposed between the promoter and the start site of the *rep* gene sequence, an AAV *rep* gene sequence, and an AAV *cap* gene sequence.

Optionally, the *rep* and/or *cap* sequences may be supplied on a vector that contains other DNA sequences that are to be introduced into the host cells. For instance, the vector may contain the rAAV construct comprising the minigene. The vector may comprise one or more of the genes encoding the helper functions, e.g., the adenoviral proteins E1, E2a, and E4ORF6, and the gene for VAI RNA.

Preferably, the promoter used in this construct may be any of the constitutive, inducible or native promoters known to one of skill in the art or as

discussed above. In one embodiment, an AAV P5 promoter sequence is employed. The selection of the AAV to provide any of these sequences does not limit the invention.

In another preferred embodiment, the promoter for *rep* is an inducible promoter, as are discussed above in connection with the transgene regulatory elements. One preferred promoter for *rep* expression is the T7 promoter. The vector comprising the *rep* gene regulated by the T7 promoter and the *cap* gene, is transfected or transformed into a cell which either constitutively or inducibly expresses the T7 polymerase. See WO 98/10088, published March 12, 1998.

The spacer is an optional element in the design of the vector. The spacer is a DNA sequence interposed between the promoter and the *rep* gene ATG start site. The spacer may have any desired design; that is, it may be a random sequence of nucleotides, or alternatively, it may encode a gene product, such as a marker gene. The spacer may contain genes which typically incorporate start/stop and polyA sites. The spacer may be a non-coding DNA sequence from a prokaryote or eukaryote, a repetitive non-coding sequence, a coding sequence without transcriptional controls or a coding sequence with transcriptional controls. Two exemplary sources of spacer sequences are the λ phage ladder sequences or yeast ladder sequences, which are available commercially, e.g., from Gibco or Invitrogen, among others. The spacer may be of any size sufficient to reduce expression of the *rep*78 and *rep*68 gene products, leaving the *rep*52, *rep*40 and *cap* gene

products expressed at normal levels. The length of the spacer may therefore range from about 10 bp to about 10.0 kbp, preferably in the range of about 100 bp to about 8.0 kbp. To reduce the possibility of recombination, the spacer is preferably less than 2 kbp in length; however, the invention is not so limited.

Although the molecule(s) providing *rep* and *cap* may exist in the host cell transiently (i.e., through transfection), it is preferred that one or both of the *rep* and *cap* proteins and the promoter(s) controlling their expression be stably expressed in the host cell, e.g., as an episome or by integration into the chromosome of the host cell. The methods employed for constructing embodiments of this invention are conventional genetic engineering or recombinant engineering techniques such as those described in the references above. While this specification provides illustrative examples of specific constructs, using the information provided herein, one of skill in the art may select and design other suitable constructs, using a choice of spacers, P5 promoters, and other elements, including at least one translational start and stop signal, and the optional addition of polyadenylation sites.

In another embodiment of this invention, the rep or cap protein may be provided stably by a host cell.

C. The Helper Functions

The packaging host cell also requires helper functions in order to package the rAAV of the invention. Optionally, these functions may be supplied

44.5 [SEQ ID NO:47], 223.10 [SEQ ID NO:48], 223.2 [SEQ ID NO:49], 223.4 [SEQ ID NO:50], 223.5 [SEQ ID NO: 51], 223.6 [SEQ ID NO: 52], 223.7 [SEQ ID NO: 53], A3.4 [SEQ ID NO: 54], A3.5 [SEQ ID NO:55], A3.7 [SEQ ID NO: 56], A3.3 [SEQ ID NO:57], 42.12 [SEQ ID NO: 58], 44.2 [SEQ ID NO: 59]. The nucleotide sequences of the signature regions of AAV10 [SEQ ID NO: 117], AAV11 [SEQ ID NO: 118] and AAV12 [SEQ ID NO:119] are provided in this figure. Critical landmarks in the structures of AAV genomes are shown. Gaps are demonstrated by dots. The 3′ ITR of AAV1 [SEQ ID NO:6] is shown in the same configuration as in the published sequences. TRS represents terminal resolution site. Notice that AAV7 is the only AAV reported that uses GTG as the initiation codon for VP3.

Figs. 2A through 2F are an alignment of the amino acid sequences of the proteins of the vp1 capsid proteins of previously published AAV serotypes 1 [SEQ ID NO:64], AAV2 [SEQ ID NO:70], AAV3 [SEQ ID NO: 71], AAV4 [SEQ ID NO:63], AAV5 [SEQ ID NO:114], and AAV6 [SEQ ID NO:65] and novel AAV sequences of the invention, including: C1 [SEQ ID NO:60], C2 [SEQ ID NO:61], C5 [SEQ ID NO:62], A3-3 [SEQ ID NO:66], A3-7 [SEQ ID NO:67], A3-4 [SEQ ID NO:68], A3-5 [SEQ ID NO: 69], 3.3b [SEQ ID NO: 62], 223.4 [SEQ ID NO: 73], 223-5 [SEQ ID NO:74], 223-10 [SEQ ID NO:75], 223-2 [SEQ ID NO:76], 223-7 [SEQ ID NO: 77], 223-6 [SEQ ID NO: 78], 44-1 [SEQ ID NO: 79], 44-5 [SEQ ID NO:80], 44-2 [SEQ ID NO:81], 42-15 [SEQ ID NO: 84], 42-8 [SEQ ID NO: 85], 42-13 [SEQ ID NO:86], 42-3A [SEQ ID NO:87], 42-4 [SEQ ID NO:88], 42-5A [SEQ ID NO:89], 42-1B [SEQ ID NO:90],

by a herpesvirus. Most desirably, the necessary helper functions are each provided from a human or non-human primate adenovirus source, such as those described above and/or are available from a variety of sources, including the American Type Culture Collection (ATCC), Manassas, VA (US). In one currently preferred embodiment, the host cell is provided with and/or contains an E1a gene product, an E1b gene product, an E2a gene product, and/or an E4 ORF6 gene product. The host cell may contain other adenoviral genes such as VAI RNA, but these genes are not required. In a preferred embodiment, no other adenovirus genes or gene functions are present in the host cell.

By "adenoviral DNA which expresses the E1a gene product", it is meant any adenovirus sequence encoding E1a or any functional E1a portion. Adenoviral DNA which expresses the E2a gene product and adenoviral DNA which expresses the E4 ORF6 gene products are defined similarly. Also included are any alleles or other modifications of the adenoviral gene or functional portion thereof. Such modifications may be deliberately introduced by resort to conventional genetic engineering or mutagenic techniques to enhance the adenoviral function in some manner, as well as naturally occurring allelic variants thereof. Such modifications and methods for manipulating DNA to achieve these adenovirus gene functions are known to those of skill in the art.

The adenovirus E1a, E1b, E2a, and/or E4ORF6 gene products, as well as any other desired helper functions, can be provided using any means that allows

their expression in a cell. Each of the sequences encoding these products may be on a separate vector, or one or more genes may be on the same vector. The vector may be any vector known in the art or disclosed above, including plasmids, cosmids and viruses. Introduction into the host cell of the vector may be achieved by any means known in the art or as disclosed above, including transfection, infection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion, among others. One or more of the adenoviral genes may be stably integrated into the genome of the host cell, stably expressed as episomes, or expressed transiently. The gene products may all be expressed transiently, on an episome or stably integrated, or some of the gene products may be expressed stably while others are expressed transiently. Furthermore, the promoters for each of the adenoviral genes may be selected independently from a constitutive promoter, an inducible promoter or a native adenoviral promoter. The promoters may be regulated by a specific physiological state of the organism or cell (i.e., by the differentiation state or in replicating or quiescent cells) or by exogenously-added factors, for example.

D. Host Cells And Packaging Cell Lines

The host cell itself may be selected from any biological organism, including prokaryotic (e.g., bacterial) cells, and eukaryotic cells, including, insect cells, yeast cells and mammalian cells.

Particularly desirable host cells are selected from among any mammalian species, including, without limitation, cells such as A549, WEHI, 3T3, 10T1/2, BHK, MDCK, COS 1, COS 7, BSC 1, BSC 40, BMT 10, VERO, WI38, HeLa, 293 cells (which express functional adenoviral E1), Saos, C2C12, L cells, HT1080, HepG2 and primary fibroblast, hepatocyte and myoblast cells derived from mammals including human, monkey, mouse, rat, rabbit, and hamster. The selection of the mammalian species providing the cells is not a limitation of this invention; nor is the type of mammalian cell, i.e., fibroblast, hepatocyte, tumor cell, etc. The most desirable cells do not carry any adenovirus gene other than E1, E2a and/or E4 ORF6; nor do they contain any other virus gene which could result in homologous recombination of a contaminating virus during the production of rAAV; and it is capable of infection or transfection of DNA and expression of the transfected DNA. In a preferred embodiment, the host cell is one that has *rep* and *cap* stably transfected in the cell.

One host cell useful in the present invention is a host cell stably transformed with the sequences encoding rep and cap, and which is transfected with the adenovirus E1, E2a, and E4ORF6 DNA and a construct carrying the minigene as described above. Stable *rep* and/or *cap* expressing cell lines, such as B-50 (PCT/US98/19463), or those described in U.S. Patent No. 5,658,785, may also be similarly employed. Another desirable host cell contains the minimum adenoviral DNA which is sufficient to express E4 ORF6. Yet other cell

lines can be constructed using the novel AAV *rep* and/or novel AAV *cap* sequences of the invention.

The preparation of a host cell according to this invention involves techniques such as assembly of selected DNA sequences. This assembly may be accomplished utilizing conventional techniques. Such techniques include cDNA and genomic cloning, which are well known and are described in Sambrook et al., cited above, use of overlapping oligonucleotide sequences of the adenovirus and AAV genomes, combined with polymerase chain reaction, synthetic methods, and any other suitable methods which provide the desired nucleotide sequence.

Introduction of the molecules (as plasmids or viruses) into the host cell may also be accomplished using techniques known to the skilled artisan and as discussed throughout the specification. In preferred embodiment, standard transfection techniques are used, e.g., $CaPO_4$ transfection or electroporation, and/or infection by hybrid adenovirus/AAV vectors into cell lines such as the human embryonic kidney cell line HEK 293 (a human kidney cell line containing functional adenovirus E1 genes which provides *trans*-acting E1 proteins).

These novel AAV-based vectors which are generated by one of skill in the art are beneficial for gene delivery to selected host cells and gene therapy patients since no neutralization antibodies to AAV7 have been found in the human population. Further, early studies show no neutralizing antibodies in cyno monkey and chimpanzee populations, and less than 15%

cross-reactivity of AAV 7 in rhesus monkeys, the species from which the serotype was isolated. One of skill in the art may readily prepare other rAAV viral vectors containing the AAV7 capsid proteins provided herein using a variety of techniques known to those of skill in the art. One may similarly prepare still other rAAV viral vectors containing AAV7 sequence and AAV capsids of another serotype. Similar advantages are conferred by the vectors based on the other novel AAV of the invention.

Thus, one of skill in the art will readily understand that the AAV7 sequences of the invention can be readily adapted for use in these and other viral vector systems for *in vitro, ex vivo or in vivo* gene delivery. Similarly, one of skill in the art can readily select other fragments of the novel AAV genome of the invention for use in a variety of rAAV and non-rAAV vector systems. Such vectors systems may include, e.g., lentiviruses, retroviruses, poxviruses, vaccinia viruses, and adenoviral systems, among others. Selection of these vector systems is not a limitation of the present invention.

Thus, the invention further provides vectors generated using the nucleic acid and amino acid sequences of the novel AAV of the invention. Such vectors are useful for a variety of purposes, including for delivery of therapeutic molecules and for use in vaccine regimens. Particularly desirable for delivery of therapeutic molecules are recombinant AAV containing capsids of the novel AAV of the invention. These, or other vector constructs containing novel AAV sequences

of the invention may be used in vaccine regimens, e.g., for co-delivery of a cytokine, or for delivery of the immunogen itself.

V. Recombinant Viruses And Uses Thereof

Using the techniques described herein, one of skill in the art may generate a rAAV having a capsid of a novel serotype of the invention, or a novel capsid containing one or more novel fragments of an AAV serotype identified by the method of the invention. In one embodiment, a full-length capsid from a single serotype, e.g., AAV7 [SEQ ID NO: 2] can be utilized. In another embodiment, a full-length capsid may be generated which contains one or more fragments of a novel serotype of the invention fused in frame with sequences from another selected AAV serotype. For example, a rAAV may contain one or more of the novel hypervariable region sequences of an AAV serotype of the invention. Alternatively, the unique AAV serotypes of the invention may be used in constructs containing other viral or non-viral sequences.

It will be readily apparent to one of skill in the art one embodiment, that certain serotypes of the invention will be particularly well suited for certain uses. For example, vectors based on AAV7 capsids of the invention are particularly well suited for use in muscle; whereas vectors based on rh.10 (44-2) capsids of the invention are particularly well suited for use in lung. Uses of such vectors are not so limited and one of skill in the art may utilize these vectors for delivery to other cell types, tissues or organs.

Further, vectors based upon other capsids of the invention may be used for delivery to these or other cells, tissues or organs.

A. Delivery of Transgene

In another aspect, the present invention provides a method for delivery of a transgene to a host which involves transfecting or infecting a selected host cell with a vector generated with the sequences of the AAV of the invention. Methods for delivery are well known to those of skill in the art and are not a limitation of the present invention.

In one desirable embodiment, the invention provides a method for AAV-mediated delivery of a transgene to a host. This method involves transfecting or infecting a selected host cell with a recombinant viral vector containing a selected transgene under the control of sequences which direct expression thereof and AAV capsid proteins.

Optionally, a sample from the host may be first assayed for the presence of antibodies to a selected AAV serotype. A variety of assay formats for detecting neutralizing antibodies are well known to those of skill in the art. The selection of such an assay is not a limitation of the present invention. See, e.g., Fisher et al, *Nature Med.*, **3**(3):306-312 (March 1997) and W. C. Manning et al, *Human Gene Therapy*, **9**:477-485 (March 1, 1998). The results of this assay may be used to determine which AAV vector containing capsid proteins of a particular serotype are preferred for

delivery, e.g., by the absence of neutralizing antibodies specific for that capsid serotype.

In one aspect of this method, the delivery of vector with a selected AAV capsid proteins may precede or follow delivery of a gene via a vector with a different serotype AAV capsid protein. Similarly, the delivery of vector with other novel AAV capsid proteins of the invention may precede or follow delivery of a gene via a vector with a different serotype AAV capsid protein. Thus, gene delivery via rAAV vectors may be used for repeat gene delivery to a selected host cell. Desirably, subsequently administered rAAV vectors carry the same transgene as the first rAAV vector, but the subsequently administered vectors contain capsid proteins of serotypes which differ from the first vector. For example, if a first vector has AAV7 capsid proteins [SEQ ID NO:2], subsequently administered vectors may have capsid proteins selected from among the other serotypes, including AAV1, AAV2, AAV3A, AAV3B, AAV4, AAV6, AAV10, AAV11, and AAV12, or any of the other novel AAV capsids identified herein including, without limitation: A3.1, H2, H6, C1, C2, C5, A3-3, A3-7, A3-4, A3-5, 3.3b, 223.4, 223-5, 223-10, 223-2, 223-7, 223-6, 44-1, 44-5, 44-2, 42-15, 42-8, 42-13, 42-3A, 42-4, 42-5A, 42-1B, 42-5B, 43-1, 43-12, 43-5, 43-21, 43-25, 43-20, 24.1, 42.2, 7.2, 27.3, 16.3, 42.10, 42-3B, 42-11, F1, F5, F3, 42-6B, and/or 42-12.

The above-described recombinant vectors may be delivered to host cells according to published methods. The rAAV, preferably suspended in a physiologically compatible carrier, may be administered to a human or

non-human mammalian patient. Suitable carriers may be readily selected by one of skill in the art in view of the indication for which the transfer virus is directed. For example, one suitable carrier includes saline, which may be formulated with a variety of buffering solutions (e.g., phosphate buffered saline). Other exemplary carriers include sterile saline, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, and water. The selection of the carrier is not a limitation of the present invention.

Optionally, the compositions of the invention may contain, in addition to the rAAV and carrier(s), other conventional pharmaceutical ingredients, such as preservatives, or chemical stabilizers. Suitable exemplary preservatives include chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, and parachlorophenol. Suitable chemical stabilizers include gelatin and albumin.

The viral vectors are administered in sufficient amounts to transfect the cells and to provide sufficient levels of gene transfer and expression to provide a therapeutic benefit without undue adverse effects, or with medically acceptable physiological effects, which can be determined by those skilled in the medical arts. Conventional and pharmaceutically acceptable routes of administration include, but are not limited to, direct delivery to the selected organ (e.g., intraportal delivery to the liver), oral, inhalation (including intranasal and intratracheal

delivery), intraocular, intravenous, intramuscular, subcutaneous, intradermal, and other parental routes of administration. Routes of administration may be combined, if desired.

Dosages of the viral vector will depend primarily on factors such as the condition being treated, the age, weight and health of the patient, and may thus vary among patients. For example, a therapeutically effective human dosage of the viral vector is generally in the range of from about 1 ml to about 100 ml of solution containing concentrations of from about 1 x 10^9 to 1 x 10^{16} genomes virus vector. A preferred human dosage may be about 1 x 10^{13} to 1 x 10^{16} AAV genomes. The dosage will be adjusted to balance the therapeutic benefit against any side effects and such dosages may vary depending upon the therapeutic application for which the recombinant vector is employed. The levels of expression of the transgene can be monitored to determine the frequency of dosage resulting in viral vectors, preferably AAV vectors containing the minigene. Optionally, dosage regimens similar to those described for therapeutic purposes may be utilized for immunization using the compositions of the invention.

Examples of therapeutic products and immunogenic products for delivery by the AAV-containing vectors of the invention are provided below. These vectors may be used for a variety of therapeutic or vaccinal regimens, as described herein. Additionally, these vectors may be delivered in combination with one or more other vectors or active ingredients in a desired therapeutic and/or vaccinal regimen.

42-5B [SEQ ID NO:91], 43-1 [SEQ ID NO: 92], 43-12 [SEQ ID NO: 93], 43-5 [SEQ ID NO:94], 43-21 [SEQ ID NO:96], 43-25 [SEQ ID NO: 97], 43-20 [SEQ ID NO:99], 24.1 [SEQ ID NO: 101], 42.2 [SEQ ID NO:102], 7.2 [SEQ ID NO: 103], 27.3 [SEQ ID NO: 104], 16.3 [SEQ ID NO: 105], 42.10 [SEQ ID NO: 106], 42-3B [SEQ ID NO: 107], 42-11 [SEQ ID NO: 108], F1 [SEQ ID NO: 109], F5 [SEQ ID NO: 110], F3 [SEQ ID NO:111], 42-6B [SEQ ID NO: 112], 42-12 [SEQ ID NO: 113]. Novel serotypes AAV8 [SEQ ID NO:95] and AAV9 [SEQ ID NO:100] are the subject of co-filed patent applications.

Figs. 3A through 3C provide the amino acid sequences of the AAV7 rep proteins [SEQ ID NO:3].

DETAILED DESCRIPTION OF THE INVENTION

In the present invention, the inventors have found a method which takes advantage of the ability of adeno-associated virus (AAV) to penetrate the nucleus, and, in the absence of a helper virus co-infection, to integrate into cellular DNA and establish a latent infection. This method utilizes a polymerase chain reaction (PCR)-based strategy for detection, identification and/or isolation of sequences of AAVs from DNAs from tissues of human and non-human primate origin as well as from other sources. Advantageously, this method is also suitable for detection, identification and/or isolation of other integrated viral and non-viral sequences, as described below.

The invention further provides nucleic acid sequences identified according to the methods of the invention. One such adeno-associated virus is of a

B. Therapeutic Transgenes

Useful therapeutic products encoded by the transgene include hormones and growth and differentiation factors including, without limitation, insulin, glucagon, growth hormone (GH), parathyroid hormone (PTH), growth hormone releasing factor (GRF), follicle stimulating hormone (FSH), luteinizing hormone (LH), human chorionic gonadotropin (hCG), vascular endothelial growth factor (VEGF), angiopoietins, angiostatin, granulocyte colony stimulating factor (GCSF), erythropoietin (EPO), connective tissue growth factor (CTGF), basic fibroblast growth factor (bFGF), acidic fibroblast growth factor (aFGF), epidermal growth factor (EGF), transforming growth factor α (TGFα), platelet-derived growth factor (PDGF), insulin growth factors I and II (IGF-I and IGF-II), any one of the transforming growth factor β superfamily, including TGF β, activins, inhibins, or any of the bone morphogenic proteins (BMP) BMPs 1-15, any one of the heregluin/neuregulin/ARIA/neu differentiation factor (NDF) family of growth factors, nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophins NT-3 and NT-4/5, ciliary neurotrophic factor (CNTF), glial cell line derived neurotrophic factor (GDNF), neurturin, agrin, any one of the family of semaphorins/collapsins, netrin-1 and netrin-2, hepatocyte growth factor (HGF), ephrins, noggin, sonic hedgehog and tyrosine hydroxylase.

Other useful transgene products include proteins that regulate the immune system including, without limitation, cytokines and lymphokines such as

thrombopoietin (TPO), interleukins (IL) IL-1 through IL-25 (including, IL-2, IL-4, IL-12, and IL-18), monocyte chemoattractant protein, leukemia inhibitory factor, granulocyte-macrophage colony stimulating factor, Fas ligand, tumor necrosis factors α and β, interferons α, β, and γ, stem cell factor, flk-2/flt3 ligand. Gene products produced by the immune system are also useful in the invention. These include, without limitations, immunoglobulins IgG, IgM, IgA, IgD and IgE, chimeric immunoglobulins, humanized antibodies, single chain antibodies, T cell receptors, chimeric T cell receptors, single chain T cell receptors, class I and class II MHC molecules, as well as engineered immunoglobulins and MHC molecules. Useful gene products also include complement regulatory proteins such as complement regulatory proteins, membrane cofactor protein (MCP), decay accelerating factor (DAF), CR1, CF2 and CD59.

Still other useful gene products include any one of the receptors for the hormones, growth factors, cytokines, lymphokines, regulatory proteins and immune system proteins. The invention encompasses receptors for cholesterol regulation, including the low density lipoprotein (LDL) receptor, high density lipoprotein (HDL) receptor, the very low density lipoprotein (VLDL) receptor, and the scavenger receptor. The invention also encompasses gene products such as members of the steroid hormone receptor superfamily including glucocorticoid receptors and estrogen receptors, Vitamin D receptors and other nuclear receptors. In addition, useful gene products include transcription

factors such as *jun*, *fos*, max, mad, serum response factor (SRF), AP-1, AP2, *myb*, MyoD and myogenin, ETS-box containing proteins, TFE3, E2F, ATF1, ATF2, ATF3, ATF4, ZF5, NFAT, CREB, HNF-4, C/EBP, SP1, CCAAT-box binding proteins, interferon regulation factor (IRF-1), Wilms tumor protein, ETS-binding protein, STAT, GATA-box binding proteins, e.g., GATA-3, and the forkhead family of winged helix proteins.

Other useful gene products include, carbamoyl synthetase I, ornithine transcarbamylase, arginosuccinate synthetase, arginosuccinate lyase, arginase, fumarylacetacetate hydrolase, phenylalanine hydroxylase, alpha-1 antitrypsin, glucose-6-phosphatase, porphobilinogen deaminase, factor VIII, factor IX, cystathione beta-synthase, branched chain ketoacid decarboxylase, albumin, isovaleryl-coA dehydrogenase, propionyl CoA carboxylase, methyl malonyl CoA mutase, glutaryl CoA dehydrogenase, insulin, beta-glucosidase, pyruvate carboxylate, hepatic phosphorylase, phosphorylase kinase, glycine decarboxylase, H-protein, T-protein, a cystic fibrosis transmembrane regulator (CFTR) sequence, and a dystrophin cDNA sequence. Still other useful gene products include enzymes such as may be useful in enzyme replacement therapy, which is useful in a variety of conditions resulting from deficient activity of enzyme. For example, enzymes that contain mannose-6-phosphate may be utilized in therapies for lysosomal storage diseases (e.g., a suitable gene includes that encoding β-glucuronidase (GUSB)).

Other useful gene products include non-naturally occurring polypeptides, such as chimeric or hybrid polypeptides having a non-naturally occurring amino acid sequence containing insertions, deletions or amino acid substitutions. For example, single-chain engineered immunoglobulins could be useful in certain immunocompromised patients. Other types of non-naturally occurring gene sequences include antisense molecules and catalytic nucleic acids, such as ribozymes, which could be used to reduce overexpression of a target.

Reduction and/or modulation of expression of a gene is particularly desirable for treatment of hyperproliferative conditions characterized by hyperproliferating cells, as are cancers and psoriasis. Target polypeptides include those polypeptides which are produced exclusively or at higher levels in hyperproliferative cells as compared to normal cells. Target antigens include polypeptides encoded by oncogenes such as myb, myc, fyn, and the translocation gene bcr/abl, ras, src, P53, neu, trk and EGRF. In addition to oncogene products as target antigens, target polypeptides for anti-cancer treatments and protective regimens include variable regions of antibodies made by B cell lymphomas and variable regions of T cell receptors of T cell lymphomas which, in some embodiments, are also used as target antigens for autoimmune disease. Other tumor-associated polypeptides can be used as target polypeptides such as polypeptides which are found at higher levels in tumor cells including the polypeptide recognized by

monoclonal antibody 17-1A and folate binding polypeptides.

Other suitable therapeutic polypeptides and proteins include those which may be useful for treating individuals suffering from autoimmune diseases and disorders by conferring a broad based protective immune response against targets that are associated with autoimmunity including cell receptors and cells which produce "self"-directed antibodies. T cell mediated autoimmune diseases include Rheumatoid arthritis (RA), multiple sclerosis (MS), Sjögren's syndrome, sarcoidosis, insulin dependent diabetes mellitus (IDDM), autoimmune thyroiditis, reactive arthritis, ankylosing spondylitis, scleroderma, polymyositis, dermatomyositis, psoriasis, vasculitis, Wegener's granulomatosis, Crohn's disease and ulcerative colitis. Each of these diseases is characterized by T cell receptors (TCRs) that bind to endogenous antigens and initiate the inflammatory cascade associated with autoimmune diseases.

C. Immunogenic Transgenes

Alternatively, or in addition, the vectors of the invention may contain AAV sequences of the invention and a transgene encoding a peptide, polypeptide or protein which induces an immune response to a selected immunogen. For example, immunogens may be selected from a variety of viral families. Example of desirable viral families against which an immune response would be desirable include, the picornavirus family, which includes the genera rhinoviruses, which are responsible

for about 50% of cases of the common cold; the genera enteroviruses, which include polioviruses, coxsackieviruses, echoviruses, and human enteroviruses such as hepatitis A virus; and the genera apthoviruses, which are responsible for foot and mouth diseases, primarily in non-human animals. Within the picornavirus family of viruses, target antigens include the VP1, VP2, VP3, VP4, and VPG. Another viral family includes the calcivirus family, which encompasses the Norwalk group of viruses, which are an important causative agent of epidemic gastroenteritis. Still another viral family desirable for use in targeting antigens for inducing immune responses in humans and non-human animals is the togavirus family, which includes the genera alphavirus, which include Sindbis viruses, RossRiver virus, and Venezuelan, Eastern & Western Equine encephalitis, and rubivirus, including Rubella virus. The flaviviridae family includes dengue, yellow fever, Japanese encephalitis, St. Louis encephalitis and tick borne encephalitis viruses. Other target antigens may be generated from the Hepatitis C or the coronavirus family, which includes a number of non-human viruses such as infectious bronchitis virus (poultry), porcine transmissible gastroenteric virus (pig), porcine hemagglutinating encephalomyelitis virus (pig), feline infectious peritonitis virus (cats), feline enteric coronavirus (cat), canine coronavirus (dog), and human respiratory coronaviruses, which may cause the common cold and/or non-A, B or C hepatitis. Within the coronavirus family, target antigens include the E1 (also called M

or matrix protein), E2 (also called S or Spike protein), E3 (also called HE or hemagglutin-elterose) glycoprotein (not present in all coronaviruses), or N (nucleocapsid). Still other antigens may be targeted against the rhabdovirus family, which includes the genera vesiculovirus (e.g., Vesicular Stomatitis Virus), and the general lyssavirus (e.g., rabies). Within the rhabdovirus family, suitable antigens may be derived from the G protein or the N protein. The family filoviridae, which includes hemorrhagic fever viruses such as Marburg and Ebola virus may be a suitable source of antigens. The paramyxovirus family includes parainfluenza Virus Type 1, parainfluenza Virus Type 3, bovine parainfluenza Virus Type 3, rubulavirus (mumps virus, parainfluenza Virus Type 2, parainfluenza virus Type 4, Newcastle disease virus (chickens), rinderpest, morbillivirus, which includes measles and canine distemper, and pneumovirus, which includes respiratory syncytial virus. The influenza virus is classified within the family orthomyxovirus and is a suitable source of antigen (e.g., the HA protein, the N1 protein). The bunyavirus family includes the genera bunyavirus (California encephalitis, La Crosse), phlebovirus (Rift Valley Fever), hantavirus (puremala is a hemahagin fever virus), nairovirus (Nairobi sheep disease) and various unassigned bungaviruses. The arenavirus family provides a source of antigens against LCM and Lassa fever virus. The reovirus family includes the genera reovirus, rotavirus (which causes acute gastroenteritis in children), orbiviruses, and cultivirus (Colorado

Tick fever, Lebombo (humans), equine encephalosis, blue tongue).

The retrovirus family includes the sub-family oncorivirinal which encompasses such human and veterinary diseases as feline leukemia virus, HTLVI and HTLVII, lentivirinal (which includes human immunodeficiency virus (HIV), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), equine infectious anemia virus, and spumavirinal). Between the HIV and SIV, many suitable antigens have been described and can readily be selected. Examples of suitable HIV and SIV antigens include, without limitation the gag, pol, Vif, Vpx, VPR, Env, Tat and Rev proteins, as well as various fragments thereof. In addition, a variety of modifications to these antigens have been described. Suitable antigens for this purpose are known to those of skill in the art. For example, one may select a sequence encoding the gag, pol, Vif, and Vpr, Env, Tat and Rev, amongst other proteins. See, e.g., the modified gag protein which is described in US Patent 5,972,596. See, also, the HIV and SIV proteins described in D.H. Barouch et al, J. Virol., 75(5):2462-2467 (March 2001), and R.R. Amara, et al, Science, 292:69-74 (6 April 2001). These proteins or subunits thereof may be delivered alone, or in combination via separate vectors or from a single vector.

The papovavirus family includes the sub-family polyomaviruses (BKU and JCU viruses) and the sub-family papillomavirus (associated with cancers or malignant progression of papilloma). The adenovirus family

includes viruses (EX, AD7, ARD, O.B.) which cause respiratory disease and/or enteritis. The parvovirus family feline parvovirus (feline enteritis), feline panleucopeniavirus, canine parvovirus, and porcine parvovirus. The herpesvirus family includes the sub-family alphaherpesvirinae, which encompasses the genera simplexvirus (HSVI, HSVII), varicellovirus (pseudorabies, varicella zoster) and the sub-family betaherpesvirinae, which includes the genera cytomegalovirus (HCMV, muromegalovirus) and the sub-family gammaherpesvirinae, which includes the genera lymphocryptovirus, EBV (Burkitts lymphoma), infectious rhinotracheitis, Marek's disease virus, and rhadinovirus. The poxvirus family includes the sub-family chordopoxvirinae, which encompasses the genera orthopoxvirus (Variola (Smallpox) and Vaccinia (Cowpox)), parapoxvirus, avipoxvirus, capripoxvirus, leporipoxvirus, suipoxvirus, and the sub-family entomopoxvirinae. The hepadnavirus family includes the Hepatitis B virus. One unclassified virus which may be suitable source of antigens is the Hepatitis delta virus. Still other viral sources may include avian infectious bursal disease virus and porcine respiratory and reproductive syndrome virus. The alphavirus family includes equine arteritis virus and various Encephalitis viruses.

The present invention may also encompass immunogens which are useful to immunize a human or non-human animal against other pathogens including bacteria, fungi, parasitic microorganisms or multicellular parasites which infect human and non-

human vertebrates, or from a cancer cell or tumor cell. Examples of bacterial pathogens include pathogenic gram-positive cocci include pneumococci; staphylococci; and streptococci. Pathogenic gram-negative cocci include meningococcus; gonococcus. Pathogenic enteric gram-negative bacilli include enterobacteriaceae; pseudomonas, acinetobacteria and eikenella; melioidosis; salmonella; shigella; haemophilus; moraxella; *H. ducreyi* (which causes chancroid); brucella; *Franisella tularensis* (which causes tularemia); yersinia (pasteurella); streptobacillus moniliformis and spirillum; Gram-positive bacilli include listeria monocytogenes; erysipelothrix rhusiopathiae; *Corynebacterium diphtheria* (diphtheria); cholera; *B. anthracis* (anthrax); donovanosis (granuloma inguinale); and bartonellosis. Diseases caused by pathogenic anaerobic bacteria include tetanus; botulism; other clostridia; tuberculosis; leprosy; and other mycobacteria. Pathogenic spirochetal diseases include syphilis; treponematoses: yaws, pinta and endemic syphilis; and leptospirosis. Other infections caused by higher pathogen bacteria and pathogenic fungi include actinomycosis; nocardiosis; cryptococcosis, blastomycosis, histoplasmosis and coccidioidomycosis; candidiasis, aspergillosis, and mucormycosis; sporotrichosis; paracoccidiodomycosis, petriellidiosis, torulopsosis, mycetoma and chromomycosis; and dermatophytosis. Rickettsial infections include Typhus fever, Rocky Mountain spotted fever, Q fever, and Rickettsialpox. Examples of mycoplasma and chlamydial infections include: mycoplasma pneumoniae;

novel serotype, termed herein serotype 7 (AAV7). Other novel adeno-associated virus serotypes provided herein include AAV10, AAV11, and AAV12. Still other novel AAV serotypes identified according to the methods of the invention are provided in the present specification. See, Figures and Sequence Listing, which is incorporated by reference.

Also provided are fragments of these AAV sequences. Among particularly desirable AAV fragments are the cap proteins, including the vp1, vp2, vp3, the hypervariable regions, the rep proteins, including rep 78, rep 68, rep 52, and rep 40, and the sequences encoding these proteins. Each of 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 cap and/or rep sequences of the invention.

As described 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

lymphogranuloma venereum; psittacosis; and perinatal chlamydial infections. Pathogenic eukaryotes encompass pathogenic protozoans and helminths and infections produced thereby include: amebiasis; malaria; leishmaniasis; trypanosomiasis; toxoplasmosis; *Pneumocystis carinii*; *Trichans*; *Toxoplasma gondii*; babesiosis; giardiasis; trichinosis; filariasis; schistosomiasis; nematodes; trematodes or flukes; and cestode (tapeworm) infections.

Many of these organisms and/or toxins produced thereby have been identified by the Centers for Disease Control [(CDC), Department of Heath and Human Services, USA], as agents which have potential for use in biological attacks. For example, some of these biological agents, include, *Bacillus anthracis* (anthrax), *Clostridium botulinum* and its toxin (botulism), *Yersinia pestis* (plague), variola major (smallpox), *Francisella tularensis* (tularemia), and viral hemorrhagic fever, all of which are currently classified as Category A agents; *Coxiella burnetti* (Q fever); Brucella species (brucellosis), *Burkholderia mallei* (glanders), *Ricinus communis* and its toxin (ricin toxin), *Clostridium perfringens* and its toxin (epsilon toxin), *Staphylococcus* species and their toxins (enterotoxin B), all of which are currently classified as Category B agents; and Nipan virus and hantaviruses, which are currently classified as Category C agents. In addition, other organisms, which are so classified or differently classified, may be identified and/or used for such a purpose in the future. It will be readily understood that the viral

vectors and other constructs described herein are useful to deliver antigens from these organisms, viruses, their toxins or other by-products, which will prevent and/or treat infection or other adverse reactions with these biological agents.

Administration of the vectors of the invention to deliver immunogens against the variable region of the T cells elicit an immune response including CTLs to eliminate those T cells. In rheumatoid arthritis (RA), several specific variable regions of T cell receptors (TCRs) which are involved in the disease have been characterized. These TCRs include V-3, V-14, V-17 and Vα-17. Thus, delivery of a nucleic acid sequence that encodes at least one of these polypeptides will elicit an immune response that will target T cells involved in RA. In multiple sclerosis (MS), several specific variable regions of TCRs which are involved in the disease have been characterized. These TCRs include V-7 and Vα-10. Thus, delivery of a nucleic acid sequence that encodes at least one of these polypeptides will elicit an immune response that will target T cells involved in MS. In scleroderma, several specific variable regions of TCRs which are involved in the disease have been characterized. These TCRs include V-6, V-8, V-14 and Vα-16, Vα-3C, Vα-7, Vα-14, Vα-15, Vα-16, Vα-28 and Vα-12. Thus, delivery of a nucleic acid molecule that encodes at least one of these polypeptides will elicit an immune response that will target T cells involved in scleroderma.

Optionally, vectors containing AAV sequences of the invention may be delivered using a prime-boost

regimen. A variety of such regimens have been described in the art and may be readily selected. See, e.g., WO 00/11140, published March 2, 2000, incorporated by reference.

Such prime-boost regimens typically involve the administration of a DNA (e.g., plasmid) based vector to prime the immune system to second, booster, administration with a traditional antigen, such as a protein or a recombinant virus carrying the sequences encoding such an antigen. In one embodiment, the invention provides a method of priming and boosting an immune response to a selected antigen by delivering a plasmid DNA vector carrying said antigen, followed by boosting, e.g., with a vector containing AAV sequences of the invention.

In one embodiment, the prime-boost regimen involves the expression of multiproteins from the prime and/or the boost vehicle. See, e.g., R.R. Amara, Science, 292:69-74 (6 April 2001) which describes a multiprotein regimen for expression of protein subunits useful for generating an immune response against HIV and SIV. For example, a DNA prime may deliver the Gag, Pol, Vif, VPX and Vpr and Env, Tat, and Rev from a single transcript. Alternatively, the SIV Gag, Pol and HIV-1 Env is delivered.

However, the prime-boost regimens are not limited to immunization for HIV or to delivery of these antigens. For example, priming may involve delivering with a first chimp vector of the invention followed by boosting with a second chimp vector, or with a composition containing the antigen itself in protein

form. In one or example, the prime-boost regimen can provide a protective immune response to the virus, bacteria or other organism from which the antigen is derived. In another desired embodiment, the prime-boost regimen provides a therapeutic effect that can be measured using convention assays for detection of the presence of the condition for which therapy is being administered.

The priming vaccine may be administered at various sites in the body in a dose dependent manner, which depends on the antigen to which the desired immune response is being targeted. The invention is not limited to the amount or situs of injection(s) or to the pharmaceutical carrier. Rather, the priming step encompasses treatment regimens which include a single dose or dosage which is administered hourly, daily, weekly or monthly, or yearly. As an example, the mammals may receive one or two priming injection containing between about 10 μg to about 50 μg of plasmid in carrier. A desirable priming amount or dosage of the priming DNA vaccine composition ranges between about 1 μg to about 10,000 μg of the DNA vaccine. Dosages may vary from about 1 μg to 1000 μg DNA per kg of subject body weight. The amount or site of injection is desirably selected based upon the identity and condition of the mammal being vaccinated.

The dosage unit of the DNA vaccine suitable for delivery of the antigen to the mammal is described herein. The DNA vaccine is prepared for administration by being suspended or dissolved in a pharmaceutically or physiologically acceptable carrier such as isotonic

saline, isotonic salts solution or other formulations which will be apparent to those skilled in such administration. The appropriate carrier will be evident to those skilled in the art and will depend in large part upon the route of administration. The compositions of the invention may be administered to a mammal according to the routes described above, in a sustained release formulation using a biodegradable biocompatible polymer, or by on-site delivery using micelles, gels and liposomes.

Optionally, the priming step of this invention also includes administering with the priming DNA vaccine composition, a suitable amount of an adjuvant, such as are defined herein.

Preferably, a boosting composition is administered about 2 to about 27 weeks after administering the priming DNA vaccine to the mammalian subject. The administration of the boosting composition is accomplished using an effective amount of a boosting vaccine composition containing or capable of delivering the same antigen as administered by the priming DNA vaccine. The boosting composition may be composed of a recombinant viral vector derived from the same viral source or from another source. Alternatively, the "boosting composition" can be a composition containing the same antigen as encoded in the priming DNA vaccine, but in the form of a protein or peptide, which composition induces an immune response in the host. In another embodiment, the boosting vaccine composition includes a composition containing a DNA sequence encoding the antigen under the control of a regulatory

sequence directing its expression in a mammalian cell, e.g., vectors such as well-known bacterial or viral vectors. The primary requirements of the boosting vaccine composition are that the antigen of the vaccine composition is the same antigen, or a cross-reactive antigen, as that encoded by the DNA vaccine.

Suitably, the vectors of the invention are also well suited for use in regimens which use non-AAV vectors as well as proteins, peptides, and/or other biologically useful therapeutic or immunogenic compounds. These regimens are particularly well suited to gene delivery for therapeutic poses and for immunization, including inducing protective immunity. Such uses will be readily apparent to one of skill in the art.

Further, a vector of the invention provides an efficient gene transfer vehicle which can deliver a selected transgene to a selected host cell *in vivo or ex vivo* even where the organism has neutralizing antibodies to one or more AAV serotypes. In one embodiment, the vector (e.g., an rAAV) and the cells are mixed *ex vivo;* the infected cells are cultured using conventional methodologies; and the transduced cells are re-infused into the patient. Further, the vectors of the invention may also be used for production of a desired gene product *in vitro.* For *in vitro* production, a desired product (e.g., a protein) may be obtained from a desired culture following transfection of host cells with a rAAV containing the molecule encoding the desired product and culturing the cell culture under conditions which permit expression.

The expressed product may then be purified and isolated, as desired. Suitable techniques for transfection, cell culturing, purification, and isolation are known to those of skill in the art.

The following examples illustrate several aspects and embodiments of the invention.

EXAMPLES

Example 1: PCR amplification, cloning and characterization of novel AAV sequences.

Tissues from nonhuman primates were screened for AAV sequences using a PCR method based on oligonucleotides to highly conserved regions of known AAVs. A stretch of AAV sequence spanning 2886 to 3143 bp of AAV1 [SEQ ID NO:6] was selected as a PCR amplicon in which a hypervariable region of the capsid protein (Cap) that is unique to each known AAV serotype, which is termed herein a "signature region," is flanked by conserved sequences. In later analysis, this signature region was shown to be located between conserved residues spanning hypervariable region 3.

An initial survey of peripheral blood of a number of nonhuman primate species revealed detectable AAV in a subset of animals from species such as rhesus macaques, cynomologous macaques, chimpanzees and baboons. However, there were no AAV sequences detected in some other species tested, including Japanese macaques, pig-tailed macaques and squirrel monkeys. A more extensive analysis of vector distribution was

conducted in tissues of rhesus monkeys of the University of Pennsylvania and Tulane colonies recovered at necropsy. This revealed AAV sequence throughout a wide array of tissues.

A. *Amplification of an AAV signature region*

DNA sequences of AAV1-6 and AAVs isolated from Goose and Duck were aligned to each other using "Clustal W" at default settings. The alignment for AAV1-6, and including the information for the novel AAV7, is provided in Fig. 1. Sequence similarities among AAVs were compared.

In the line of study, a 257 bp region spanning 2886 bp to 3143 bp of AAV 1 [SEQ ID NO: 6], and the corresponding region in the genomes of AAV 2-6 genomes [See, Fig. 1], was identified by the inventors. This region is located with the AAV capsid gene and has highly conserved sequences among at both 5′ and 3′ ends and is relatively variable sequence in the middle. In addition, this region contains a DraIII restriction enzyme site (CACCACGTC, SEQ ID NO:15) . The inventors have found that this region serves as specific signature for each known type of AAV DNA. In other words, following PCR reactions, digestion with endonucleases that are specific to each known serotypes and gel electrophoresis analysis, this regions can be used to definitively identify amplified DNA as being from serotype 1, 2, 3, 4, 5, 6, or another serotype.

The primers were designed, validated and PCR conditions optimized with AAV1, 2 and 5 DNA controls. The primers were based upon the sequences of AAV2: 5' primer, 1S: bp 2867-2891 of AAV2 (SEQ ID NO:7) and 3'

primer, 18as, bp 3095-3121 of AAV2 (SEQ ID NO:7).

Cellular DNAs from different tissues including blood, brain, liver, lung, testis, etc. of different rhesus monkeys were studied utilizing the strategy described above. The results revealed that DNAs from different tissues of these monkeys gave rise to strong PCR amplifications. Further restriction analyses of PCR products indicated that they were amplified from AAV sequences different from any published AAV sequences.

PCR products (about 255 bp in size) from DNAs of a variety of monkey tissues have been cloned and sequenced. Bioinformatics study of these novel AAV sequences indicated that they are novel AAV sequences of capsid gene and distinct from each other. Fig. 1 includes in the alignment the novel AAV signature regions for AAV10-12 [SEQ ID NO:117, 118 and 119, respectively]. Multiple sequence alignment analysis was performed using the Clustal W (1.81) program. The percentage of sequence identity between the signature regions of AAV 1-7 and AAV 10-12 genomes is provided below.

Table 1. Sequences for Analysis

Sequence #	AAV Serotype	Size (bp)
1	AAV1	258
2	AAV2	255
3	AAV3	255
4	AAV4	246
5	AAV5	258
6	AAV6	258
7	AAV7	258
10	AAV10	255
11	AAV11	258
12	AAV12	255

Table 3. Pairwise Alignment (Percentage of Identity)

	AAV 2	AAV 3	AAV 4	AAV 5	AAV 6	AAV 7	AAV10	AAV11	AAV12
AAV1	90	90	81	76	97	91	93	94	93
AAV2		93	79	78	90	90	93	93	92
AAV3			80	76	90	92	92	92	92
AAV4				76	81	84	82	81	79
AAV5					75	78	79	79	76
AAV6						91	92	94	94
AAV7							94	92	92
AAV10								95	93
AAV11									94

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. 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.

The term "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. Preferably, the homology is over full-length sequence, 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.

The term "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

Over 300 clones containing novel AAV serotype sequences that span the selected 257 bp region were isolated and sequenced. Bioinformatics analysis of these 300+ clones suggests that this 257 bp region is critical in serving as a good land marker or signature sequence for quick isolation and identification of novel AAV serotype.

B. *Use of the signature region for PCR amplification.*

The 257 bp signature region was used as a PCR anchor to extend PCR amplifications to 5′ of the genome to cover the junction region of rep and cap genes (1398 bp - 3143 bp, SEQ ID NO:6) and 3′ of the genome to obtain the entire cap gene sequence (2866 bp - 4600 bp, SEQ ID NO:6). PCR amplifications were carried out using the standard conditions, including denaturing at 95°C for 0.5-1 min, annealing at 60-65°C for 0.5-1 min and extension at 72° C for 1 min per kb with a total number of amplification cycles ranging from 28 to 42.

Using the aligned sequences as described in "A", two other relative conserved regions were identified in the sequence located in 3′ end of rep genes and 5′ to the 257 bp region and in the sequence down stream of the 257 bp fragment but before the AAV′ 3 ITR. Two sets of new primers were designed and PCR conditions optimized for recovery of entire capsid and a part of rep sequences of novel AAV serotypes. More specifically, for the 5' amplification, the 5' primer, AV1Ns, was GCTGCGTCAACTGGACCAATGAGAAC [nt 1398-1423 of

AAV1, SEQ ID NO:6] and the 3' primer was 18as, identified above. For the 3' amplification, the 5' primer was 1s, identified above, and the 3' primer was AV2Las, TCGTTTCAGTTGAACTTTGGTCTCTGCG [nt 4435-4462 of AAV2, SEQ ID NO:7].

In these PCR amplifications, the 257 bp region was used as a PCR anchor and land marker to generate overlapping fragments to construct a complete capsid gene by fusion at the DraIII site in the signature region following amplification of the 5' and 3' extension fragments obtained as described herein. More particularly, to generate the intact AAV7 cap gene, the three amplification products (a) the sequences of the signature region; (b) the sequences of the 5' extension; and (c) the sequences of the 3' extension were cloned into a pCR4-Topo [Invitrogen] plasmid backbone according to manufacturer's instructions. Thereafter, the plasmids were digested with DraIII and recombined to form an intact cap gene.

In this line of work, about 80 % of capsid sequences of AAV7 and AAV 8 were isolated and analyzed. Another novel serotype, AAV9, was also discovered from Monkey #2.

Using the PCR conditions described above, the remaining portion of the rep gene sequence for AAV7 is isolated and cloned using the primers that amplify 108 bp to 1461 bp of AAV genome (calculated based on the numbering of AAV2, SEQ ID NO:7). This clone is sequenced for construction of a complete AAV7 genome without ITRs.

C. Direct Amplification of 3.1 kb Cap fragment

To directly amplify a 3.1 kb full-length Cap fragment from NHP tissue and blood DNAs, two other highly conserved regions were identified in AAV genomes for use in PCR amplification of large fragments. A primer within a conserved region located in the middle of the rep gene was selected (AV1ns: 5' GCTGCGTCAACTGGACCAATGAGAAC 3', nt 1398-1423 of SEQ ID NO:6) in combination with the 3' primer located in another conserved region downstream of the Cap gene (AV2cas: 5' CGCAGAGACCAAAGTTCAACTGAAACGA 3', SEQ ID NO:7) for amplification of full-length cap fragments. The PCR products were Topo-cloned according to manufacturer's directions (Invitrogen) and sequence analysis was performed by Qiagengenomics (Qiagengenomics, Seattle, WA) with an accuracy of ≥ 99.9%. A total of 50 capsid clones were isolated and characterized. Among them, 37 clones were derived from Rhesus macaque tissues (rh.1 - rh.37), 6 clones from cynomologous macaques (cy.1 - cy.6), 2 clones from Baboons (bb.1 and bb.2) and 5 clones from Chimps (ch.1 - ch.5).

To rule out the possibility that sequence diversity within the novel AAV family was not an artifact of the PCR, such as PCR-mediated gene splicing by overlap extension between different partial DNA templates with homologous sequences, or the result of recombination process in bacteria, a series of experiments were performed under identical conditions for VP1 amplification using total cellular DNAs. First, intact AAV7 and AAV8 plasmids were mixed at an equal

molar ratio followed by serial dilutions. The serially diluted mixtures were used as templates for PCR amplification of 3.1 kb VP1 fragments using universal primers and identical PCR conditions to that were used for DNA amplifications to see whether any hybrid PCR products were generated. The mixture was transformed into bacteria and isolated transformants to look for hybrid clones possibly derived from recombination process in bacterial cells. In a different experiment, we restricted AAV7 and AAV8 plasmids with Msp I, Ava I and HaeI, all of which cut both genomes multiple times at different positions, mixed the digestions in different combinations and used them for PCR amplification of VP1 fragments under the same conditions to test whether any PCR products could be generated through overlap sequence extension of partial AAV sequences. In another experiment, a mixture of gel purified 5' 1.5 kb AAV7 VP1 fragment and 3' 1.7 kb AAV8 VP1 fragment with overlap in the signature region was serially diluted and used for PCR amplification in the presence and absence of 200 ng cellular DNA extracted from a monkey cell line that was free of AAV sequences by TaqMan analysis. None of these experiments demonstrated efficient PCR-mediated overlap sequence production under the conditions of the genomic DNA Cap amplification (data not shown). As a further confirmation, 3 pairs of primers were designed, which were located at different HVRs, and were sequence specific to the variants of clone 42s from Rhesus macaque F953, in different combinations to amplify shorter fragments from mesenteric lymph node (MLN) DNA

from F953 from which clone 42s were isolated. All sequence variations identified in full-length Cap clones were found in these short fragments (data not shown).

Example 2: Adeno-Associated Viruses Undergo Substantial Evolution in Primates During Natural Infections

Sequence analysis of selected AAV isolates revealed divergence throughout the genome that is most concentrated in hypervariable regions of the capsid proteins. Epidemiologic data indicate that all known serotypes are endemic to primates, although isolation of clinical isolates has been restricted to AAV2 and AAV3 from anal and throat swabs of human infants and AAV5 from a human condylomatous wart. No known clinical sequalae have been associated with AAV infection.

In an attempt to better understand the biology of AAV, nonhuman primates were used as models to characterize the sequlae of natural infections. Tissues from nonhuman primates were screened for AAV sequences using the PCR method of the invention based on oligonucleotides to highly conserved regions of known AAVs (see Example 1). A stretch of AAV sequence spanning 2886 to 3143 bp of AAV1 [SEQ ID NO:6] was selected as a PCR amplicon in which conserved sequences are flanked by a hypervariable region that is unique to each known AAV serotype, termed herein a "signature region."

An initial survey of peripheral blood of a number of nonhuman primate species including rhesus monkeys,

cynomologous monkeys, chimpanzees, and baboons revealed detectable AAV in a subset of animals from all species. A more extensive analysis of vector distribution was conducted in tissues of rhesus monkeys of the University of Pennsylvania and Tulane colonies recovered at necropsy. This revealed AAV sequence throughout a wide array of tissues.

The amplified signature sequences were subcloned into plasmids and individual transformants were subjected to sequence analysis. This revealed substantial variation in nucleotide sequence of clones derived from different animals. Variation in the signature sequence was also noted in clones obtained within individual animals. Tissues harvested from two animals in which unique signature sequences were identified (i.e., colon from 98E044 and heart from 98E056) were further characterized by expanding the sequence amplified by PCR using oligonucleotides to highly conserved sequences. In this way, complete proviral structures were reconstructed for viral genomes from both tissues as described herein. These proviruses differ from the other known AAVs with the greatest sequence divergence noted in regions of the Cap gene.

Additional experiments were performed to confirm that AAV sequences resident to the nonhuman primate tissue represented proviral genomes of infectious virus that is capable of being rescued and form virions. Genomic DNA from liver tissue of animal 98E056, from which AAV8 signature sequence was detected, was digested with an endonuclease that does not have a site

within the AAV sequence and transfected into 293 cells with a plasmid containing an E1 deleted genome of human adenovirus serotype 5 as a source of helper functions. The resulting lysate was passaged on 293 cells once and the lysate was recovered and analyzed for the presence of AAV Cap proteins using a broadly reacting polyclonal antibody to Cap proteins and for the presence and abundance of DNA sequences from the PCR amplified AAV provirus from which AAV8 was derived. Transfection of endonuclease restricted heart DNA and the adenovirus helper plasmid yielded high quantities of AAV8 virus as demonstrated by the detection of Cap proteins by Western blot analysis and the presence of 10^4 AAV8 vector genomes per 293 cell. Lysates were generated from a large-scale preparation and the AAV was purified by cesium sedimentation. The purified preparation demonstrated 26 nm icosohedral structures that look identical to those of AAV serotype 2. Transfection with the adenovirus helper alone did not yield AAV proteins or genomes, ruling out contamination as a source of the rescued AAV.

To further characterize the inter and intra animal variation of AAV signature sequence, selected tissues were subjected to extended PCR to amplify entire Cap open reading frames.

The resulting fragments were cloned into bacterial plasmids and individual transformants were isolated and fully sequenced. This analysis involved mesenteric lymph nodes from three rhesus monkeys (Tulane/V223 - 6 clones; Tulane/T612 - 7 clones; Tulane/F953 - 14 clones), liver from two rhesus monkeys (Tulane/V251 - 3

clones; Penn/00E033 - 3 clones), spleen from one rhesus monkey (Penn/97E043 - 3 clones), heart from one rhesus monkey (IHGT/98E046- 1 clone) and peripheral blood from one chimpanzee (New Iberia/X133 - 5 clones), six cynomologous macaques (Charles River/A1378, A3099, A3388, A3442, A2821, A3242 - 6 clones total) and one Baboon (SFRB/8644 - 2 clones). Of the 50 clones that were sequenced from 15 different animals, 30 were considered non-redundant based on the finding of at least 7 amino acid differences from one another. The non-redundant VP1 clones are numbered sequentially as they were isolated, with a prefix indicating the species of non-human primate from which they were derived. The structural relationships between these 30 non-redundant clones and the previously described 8 AAV serotypes were determined using the SplitsTree program [Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68-73 (1998)] with implementation of the method of split decomposition. The analysis depicts homoplasy between a set of sequences in a tree-like network rather than a bifurcating tree. The advantage is to enable detection of groupings that are the result of convergence and to exhibit phylogenetic relationships even when they are distorted by parallel events. Extensive phylogenetic research will be required in order to elucidate the AAV evolution, whereas the intention here only is to group the different clones as to their sequence similarity.

To confirm that the novel VP1 sequences were derived from infectious viral genomes, cellular DNA from tissues with high abundance of viral DNA was

restricted with an endonuclease that should not cleave within AAV and transfected into 293 cells, followed by infection with adenovirus. This resulted in rescue and amplification of AAV genomes from DNA of tissues from two different animals (data not shown).

VP1 sequences of the novel AAVs were further characterized with respect to the nature and location of amino acid sequence variation. All 30 VP1 clones that were shown to differ from one another by greater than 1% amino acid sequence were aligned and scored for variation at each residue. An algorithm developed to determine areas of sequence divergence yielded 12 hypervariable regions (HVR) of which 5 overlap or are part of the 4 previously described variable regions [Kotin, cited above; Rutledge, cited above]. The three-fold-proximal peaks contain most of the variability (HVR5-10). Interestingly the loops located at the 2 and 5 fold axis show intense variation as well. The HVRs 1 and 2 occur in the N-terminal portion of the capsid protein that is not resolved in the X-ray structure suggesting that the N-terminus of the VP1 protein is exposed on the surface of the virion.

Real-time PCR was used to quantify AAV sequences from tissues of 21 rhesus monkeys using primers and probes to highly conserved regions of Rep (one set) and Cap (two sets) of known AAVs. Each data point represents analysis from tissue DNA from an individual animal. This confirmed the wide distribution of AAV sequences, although the quantitative distribution differed between individual animals. The source of animals and previous history or treatments did not

appear to influence distribution of AAV sequences in rhesus macaques. The three different sets of primers and probes used to quantify AAV yielded consistent results. The highest levels of AAV were found consistently in mesenteric lymph nodes at an average of 0.01 copies per diploid genome for 13 animals that were positive. Liver and spleen also contained high abundance of virus DNA. There were examples of very high AAV, such as in heart of rhesus macaque 98E056, spleen of rhesus macaque 97E043 and liver of rhesus macaque RQ4407, which demonstrated 1.5, 3 and 20 copies of AAV sequence per diploid genome respectively. Relatively low levels of virus DNA were noted in peripheral blood mononuclear cells, suggesting the data in tissue are not due to resident blood components (data not shown). It should be noted that this method would not necessarily capture all AAVs resident to the nonhuman primates since detection requires high homology to both the oligonucleotides and the real time PCR probe. Tissues from animals with high abundance AAV DNA was further analyzed for the molecular state of the DNA, by DNA hybridization techniques, and its cellular distribution, by *in situ* hybridization.

The kind of sequence variation revealed in AAV proviral fragments isolated from different animals and within tissues of the same animals is reminiscent of the evolution that occurs for many RNA viruses during pandemics or even within the infection of an individual. In some situations the notion of a wild-type virus has been replaced by the existence of swarms of quasispecies that evolve as a result of rapid

What is claimed is:

1. An adeno-associated virus (AAV) comprising an AAV capsid comprising an AAVrh10 vp1, comprising amino acids 1 to 738 of SEQ ID NO:81 or a sequence which is at least 95% identical therewith, a minigene having AAV inverted terminal repeats (ITRs), and a heterologous gene operably linked to regulatory sequences which direct its expression in a host cell.

2. The AAV according to claim 1, wherein the amino acid sequence is at least 99% identical to amino acids 1 to 738 of SEQ ID NO: 81.

3. The AAV according to claim 1 or claim 2, wherein the AAVrh10 vp1 protein comprises the sequence of amino acids 1 to 738 of SEQ ID NO: 81.

4. The AAV according to claim 1 wherein the ITRs are from AAV2.

5. An adeno-associated virus (AAV) comprising an AAV capsid comprising at least an AAVrh10 vp3 comprising the amino acid sequence of amino acids 204 to 738 of SEQ ID NO:81 or a sequence at least 95% identical thereto and a minigene having AAV inverted terminal repeats (ITRs), and a heterologous gene operably linked to regulatory sequences which direct its expression in a host cell.

12. A molecule according to claim 9, wherein said molecule further comprises a functional AAV *rep* gene.

13. A method of generating a recombinant adeno-associated virus (AAV) comprising an AAV serotype rh10 capsid comprising the steps of culturing a host cell containing: (a) a molecule according to claim 9 which encodes an adeno-associated virus capsid; (b) a functional rep gene; (c) a minigene comprising AAV inverted terminal repeats (ITRs) and a transgene; and (d) sufficient helper functions to permit packaging of the minigene into the AAV capsid protein.

14. An *in vitro* host cell containing an adeno-associated virus according to claim 1.

15. An *in vitro* host cell containing a molecule according to claim 9.

16. The composition according to claim 7, wherein the heterologous gene encodes a Factor VIII.

17. The composition according to claim 7, wherein the heterologous gene encodes a Factor IX.

18. The composition according to claim 7, wherein the heterologous gene encodes alpha-1 antitrpysin (A1AT).

P1

FIG. 1A

```
          1                                                  50
   42_2   .......... .......... .......... .......... ..........
   42_8   .......... .......... .......... .......... ..........
  42_15   .......... .......... .......... .......... ..........
  42_5b   .......... .......... .......... .......... ..........
  42_1b   .......... .......... .......... .......... ..........
  42_13   .......... .......... .......... .......... ..........
  42_3a   .......... .......... .......... .......... ..........
   42_4   .......... .......... .......... .......... ..........
  42_5a   .......... .......... .......... .......... ..........
  42_10   .......... .......... .......... .......... ..........
  42_3b   .......... .......... .......... .......... ..........
  42_11   .......... .......... .......... .......... ..........
  42_6b   .......... .......... .......... .......... ..........
   43_1   .......... .......... .......... .......... ..........
   43_5   .......... .......... .......... .......... ..........
  43_12   .......... .......... .......... .......... ..........
  43_20   .......... .......... .......... .......... ..........
  43_21   .......... .......... .......... .......... ..........
  43_23   .......... .......... .......... .......... ..........
  43_25   .......... .......... .......... .......... ..........
   44_1   .......... .......... .......... .......... ..........
   44_5   .......... .......... .......... .......... ..........
 223_10   .......... .......... .......... .......... ..........
  223_2   .......... .......... .......... .......... ..........
  223_4   .......... .......... .......... .......... ..........
  223_5   .......... .......... .......... .......... ..........
  223_6   .......... .......... .......... .......... ..........
  223_7   .......... .......... .......... .......... ..........
   A3_4   .......... .......... .......... .......... ..........
   A3_5   .......... .......... .......... .......... ..........
   A3_7   .......... .......... .......... .......... ..........
   A3_3   .......... .......... .......... .......... ..........
  42_12   .......... .......... .......... .......... ..........
   AAV1   TTGCCCACTC CCTCTCTGCG CGCTCGCTCG CTCGGTGGGG CCTGCGGACC
   AAV2   TTGGCCACTC CCTCTCTGCG CGCTCGCTCG CTCACTGAGG CCGGGCGACC
   AAV3   TTGGCCACTC CCTCTATGCG CACTCGCTCG CTCGGTGGGG CCTGGCGACC
   AAV8   .......... .......... .......... .......... ..........
   AAV9   .......... .......... .......... .......... ..........
   AAV7   TTGGCCACTC CCTCTATGCG CGCTCGCTCG CTCGGTGGGG CCTGCGGACC
   44_2   .......... .......... .......... .......... ..........
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicants
SAPALO VELEZ BUNDANG & BULILAN

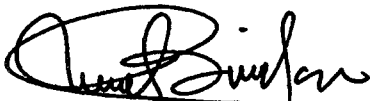
By: 
Attorneys

Fig. 1I

```
          401                                                        450
  42_2    .......... .......... .......... .......... ..........
  42_8    .......... .......... .......... .......... ..........
 42_15    .......... .......... .......... .......... ..........
 42_5b    .......... .......... .......... .......... ..........
 42_1b    .......... .......... .......... .......... ..........
 42_13    .......... .......... .......... .......... ..........
 42_3a    .......... .......... .......... .......... ..........
  42_4    .......... .......... .......... .......... ..........
 42_5a    .......... .......... .......... .......... ..........
 42_10    .......... .......... .......... .......... ..........
 42_3b    .......... .......... .......... .......... ..........
 42_11    .......... .......... .......... .......... ..........
 42_6b    .......... .......... .......... .......... ..........
  43_1    .......... .......... .......... .......... ..........
  43_5    .......... .......... .......... .......... ..........
 43_12    .......... .......... .......... .......... ..........
 43_20    .......... .......... .......... .......... ..........
 43_21    .......... .......... .......... .......... ..........
 43_23    .......... .......... .......... .......... ..........
 43_25    .......... .......... .......... .......... ..........
  44_1    .......... .......... .......... .......... ..........
  44_5    .......... .......... .......... .......... ..........
223_10    .......... .......... .......... .......... ..........
 223_2    .......... .......... .......... .......... ..........
 223_4    .......... .......... .......... .......... ..........
 223_5    .......... .......... .......... .......... ..........
 223_6    .......... .......... .......... .......... ..........
 223_7    .......... .......... .......... .......... ..........
  A3_4    .......... .......... .......... .......... ..........
  A3_5    .......... .......... .......... .......... ..........
  A3_7    .......... .......... .......... .......... ..........
  A3_3    .......... .......... .......... .......... ..........
 42_12    .......... .......... .......... .......... ..........
  AAV1    TTTCTGACTC GTTTGTGAGC TGGGTGGCCG AGAAGGAATG GGAGCTGCCC
  AAV2    TTTCTGACAG CTTTGTGAAC TGGGTGGCCG AGAAGGAATG GGAGTTGCCG
  AAV3    TTTCTAACTC GTTTGTTAAC TGGGTGGCCG AGAAGGAATG GGACGTGCCG
  AAV8    TTTCTGACTC GTTTGTGAAC TGGGTGGCCG AGAAGGAATG GGAGCTGCCC
  AAV9    TTTCTGACTC TTTTGTGAAC TGGGTGGCCG AGAAGGAATG GGAGCTGCCC
  AAV7    TTTCTGACTC GTTTGTGAAC TGGGTGGCCG AGAAGGAATG GGAGCTGCCC
  44_2    .......... .......... .......... .......... ..........
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 1S

```
          901                                                        950
  42_2    .......... .......... .......... .......... ..........
  42_8    .......... .......... .......... .......... ..........
 42_15    .......... .......... .......... .......... ..........
 42_5b    .......... .......... .......... .......... ..........
 42_1b    .......... .......... .......... .......... ..........
 42_13    .......... .......... .......... .......... ..........
 42_3a    .......... .......... .......... .......... ..........
  42_4    .......... .......... .......... .......... ..........
 42_5a    .......... .......... .......... .......... ..........
 42_10    .......... .......... .......... .......... ..........
 42_3b    .......... .......... .......... .......... ..........
 42_11    .......... .......... .......... .......... ..........
 42_6b    .......... .......... .......... .......... ..........
  43_1    .......... .......... .......... .......... ..........
  43_5    .......... .......... .......... .......... ..........
 43_12    .......... .......... .......... .......... ..........
 43_20    .......... .......... .......... .......... ..........
 43_21    .......... .......... .......... .......... ..........
 43_23    .......... .......... .......... .......... ..........
 43_25    .......... .......... .......... .......... ..........
  44_1    .......... .......... .......... .......... ..........
  44_5    .......... .......... .......... .......... ..........
223_10    .......... .......... .......... .......... ..........
 223_2    .......... .......... .......... .......... ..........
 223_4    .......... .......... .......... .......... ..........
 223_5    .......... .......... .......... .......... ..........
 223_6    .......... .......... .......... .......... ..........
 223_7    .......... .......... .......... .......... ..........
  A3_4    .......... .......... .......... .......... ..........
  A3_5    .......... .......... .......... .......... ..........
  A3_7    .......... .......... .......... .......... ..........
  A3_3    .......... .......... .......... .......... ..........
 42_12    .......... .......... .......... .......... ..........
  AAV1    CGGCTCGTGG CGCAGCACCT GACCCACGTC AGCCAGACCC AGGAGCAGAA
  AAV2    CGGTTGGTGG CGCAGCATCT GACGCACGTG TCGCAGACGC AGGAGCAGAA
  AAV3    CGGCTGGTGG CGCAGCATCT GACGCACGTG TCGCAGACGC AGGAGCAGAA
  AAV8    CGGCTCGTGG CGCAGCACCT GACCCACGTC AGCCAGACGC AGGAGCAGAA
  AAV9    CGGCTCGTGG CGCAGCACCT GACCCACGTC AGCCAGACGC AGGAGCAGAA
  AAV7    CGGCTCGTGG CGCAGCACCT GACCCACGTC AGCCAGACGC AGGAGCAGAA
  44_2    .......... .......... .......... .......... ..........
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 1AC

```
         1401                                                      1450
 42_2    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 42_8    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
42_15    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
42_5b    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
42_1b    .......... .......... .......... .......... ..........
42_13    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
42_3a    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 42_4    .......... .......... .......... .......... ..........
42_5a    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
42_10    .......... .......... .......... .......... ..........
42_3b    .......... .......... .......... .......... ..........
42_11    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
42_6b    .GTCTTCCGC CCAGATCGAT CCCACCCCCG TGATCGTCAC TTCCAACACC
 43_1    .CTACGGCTG CATCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 43_5    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
43_12    .....GGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
43_20    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
43_21    .....GGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
43_23    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
43_25    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 44_1    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 44_5    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
223_10   .......... .......... .......... .......... ..........
223_2    .......... .......... .......... .......... ..........
223_4    .......... .......... .......... .......... ..........
223_5    .......... .......... .......... .......... ..........
223_6    .......... .......... .......... .......... ..........
223_7    .......... .......... .......... .......... ..........
 A3_4    TCTACGGCTG CGTCAACTGG ACCAATGAAA ACTTTCCCTT CAACGATTGC
 A3_5    TCTACGGCTG CGTCAACTGG ACCAATGAAA ACTTTCCCTT CAACGATTGC
 A3_7    TCTACGGCTG CGTCAACTGG ACCAATGAAA ACTTTCCCTT CAACGATTGC
 A3_3    TCTACGGCTG CGTCAACTGG ACCAATGAAA ACTTTCCCTT CAACGATTGC
42_12    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 AAV1    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAATGATTGC
 AAV2    .CTACGGGTG CGTAAACTGG ACCAATGAGA ACTTTCCCTT CAACGACTGT
 AAV3    .CTACGGCTG CGTAAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 AAV8    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAATGATTGC
 AAV9    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 AAV7    .CTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
 44_2    TCTACGGCTG CGTCAACTGG ACCAATGAGA ACTTTCCCTT CAACGATTGC
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 1AM

```
         1901                                                      1950
  42_2   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  42_8   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 42_15   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 42_5b   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 42_1b   .......... .......... .......... .......... ..........
 42_13   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 42_3a   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  42_4   .......... .......... .......... .......... ..........
 42_5a   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 42_10   .......... .......... .......... .......... ..........
 42_3b   .......... .......... .......... .......... ..........
 42_11   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 42_6b   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAGTGTTC
  43_1   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  43_5   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 43_12   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 43_20   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 43_21   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 43_23   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
 43_25   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  44_1   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  44_5   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
223_10   .......... .......... .......... .......... ..........
 223_2   .......... .......... .......... .......... ..........
 223_4   .......... .......... .......... .......... ..........
 223_5   .......... .......... .......... .......... ..........
 223_6   .......... .......... .......... .......... ..........
 223_7   .......... .......... .......... .......... ..........
  A3_4   GCGGA...AG CTTCGATAAA CTACGCGGGC AGGTACCAAA ACAAATGTTC
  A3_5   GCGGA...AG CTTCGATAAA CTACGCGGAC AGGTACCAAA ACAAATGTTC
  A3_7   GCGGA...AG CTTCGATAAA CTACGCGGAC AGGTACCAAA ACAAATGTTC
  A3_3   GCGGA...AG CTTCGATAAA CTACGCGGAC AGGTACCAAA ACAAATGTTC
 42_12   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  AAV1   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  AAV2   GCGGA...AG CTTCGATCAA CTACGCAGAC AGGTACCAAA ACAAATGTTC
  AAV3   GCGGA...AG CACCGGCGGA CTACGCGGAC AGGTACCAAA ACAAATGTTC
  AAV8   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  AAV9   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  AAV7   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
  44_2   GCGGAAGGAG CTCCGGTGGA CTTTGCCGAC AGGTACCAAA ACAAATGTTC
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 1AW

```
          2401                                                      2450
  42_2    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
  42_8    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 42_15    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 42_5b    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_1b    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_13    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 42_3a    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  42_4    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_5a    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_10    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_3b    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_11    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGCG GCGGACGCAG
 42_6b    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
  43_1    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  43_5    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 43_12    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 43_20    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 43_21    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 43_23    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
 43_25    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  44_1    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  44_5    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
223_10    .......... .......... .......... .......... ..........
223_2     .......... .......... .......... .......... ..........
223_4     .......... .......... .......... .......... ..........
223_5     .......... .......... .......... .......... ..........
223_6     .......... .......... .......... .......... ..........
223_7     .......... .......... .......... .......... ..........
  A3_4    ACCCTTCAAC GGACTCGACA AAGGAGAGCC GGTCAACGAG GCAGACGCCG
  A3_5    ACCCTTCAAC GGACTCGACA AAGGAGAGCC GGTCAACGAG GCAGACGCCG
  A3_7    ACCCTTCAAC GGACTCGACA AAGGAGAGCC GGTCAACGAG GCAGACGCCG
  A3_3    ACCCTTCAAC GGACTCGACA AAGGAGAGCC GGTCAACGAG GCAGACGCCG
 42_12    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
  AAV1    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  AAV2    ACCCTTCAAC GGACTCGACA AGGGAGAGCC GGTCAACGAG GCAGACGCCG
  AAV3    ACCCGGTAAC GGACTCGACA AAGGAGAGCC GGTCAACGAG GCGGACGCGG
  AAV8    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  AAV9    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  AAV7    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
  44_2    ACCCTTCAAC GGACTCGACA AGGGGGAGCC CGTCAACGCG GCGGACGCAG
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

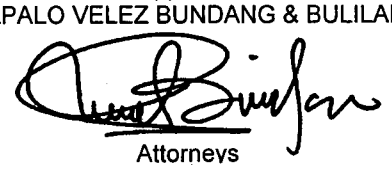
By:

Attorneys

Fig. 1AAG

```
          2901                                                    2950
  42_2    GGAGTGGGTA ATGCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  42_8    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_15    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_5b    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_1b    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_13    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_3a    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATAGCTGGG
  42_4    GGAGTGGGTA ATGCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_5a    GGAGTGGGTA ATGCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_10    GGAGTGGGTA ATGCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_3b    GGAGTGGGTA ATGCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_11    GGAGTGGGTA ATGCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 42_6b    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  43_1    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  43_5    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 43_12    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 43_20    GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 43_21    GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 43_23    GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 43_25    GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  44_1    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  44_5    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
223_10    GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
223_2     GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
223_4     GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CACGGCTGGG
223_5     GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CACGGCTGGG
223_6     GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
223_7     GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 A3_4     GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGATGGG
 A3_5     GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGATGGG
 A3_7     GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGATGGG
 A3_3     GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGATGGG
 42_12    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  AAV1    GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  AAV2    GGAGTGGGTA ATTCCTCCGG AAATTGGCAT TGCGATTCCA CATGGATGGG
  AAV3    GGAGTGGGTA ATTCCTCAGG AAATTGGCAT TGCGATTCCC AATGGCTGGG
  AAV8    GGAGTGGGTA GTTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  AAV9    GGAGTGGGTA ATTCCTCGGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  AAV7    GGAGTGGGTA ATGCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 AAV10          GGTA ATTCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 AAV11          GGTA ATTCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
 AAV12          GGTA ATTCCTCCGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
  44_2    GGAGTGGGTA GTTCCTCAGG AAATTGGCAT TGCGATTCCA CATGGCTGGG
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By: 

Attorneys

Fig. 1AAQ

```
          3401                                                    3450
  42_2    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
  42_8    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_15    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_5b    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_1b    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_13    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_3a    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
  42_4    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_5a    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_10    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_3b    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_11    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 42_6b    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
  43_1    TCAGGCTGTG GGCCGTTCCT CCTTCTACTG CCTGGAATAC TTCCCTTCTC
  43_5    TCAGGCTGTG GGCCGTTCCT CCTTCTACTG CCTGGAATAC TTCCCTTCTC
 43_12    TCAGGCTGTG GGCCGTTCCT CCTTCTACTG CCTGGAATAC TTCCCTTCTC
 43_20    CCAAGCCCTG GGACGTTCCT CCTTCTACTG TCTGGAGTAT TTCCCATCGC
 43_21    CCAAGCCCTG GGACGTTCCT CCTTCTACTG TCTGGAGTAT TTCCCATCGC
 43_23    CCAAGCCCTG GGACGTTCCT CCTTCTACTG TCTGGAGTAT TTCCCATCGC
 43_25    CCAAGCCCTG GGACGTTCCT CCTTCTACTG TCTGGAGTAT TTCCCATCGC
  44_1    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
  44_5    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
223_10    CCAATCGGTA GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 223_2    CCAATCGGTA GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 223_4    CCAATCGGTA GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 223_5    CCAATCGGTA GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 223_6    CCAATCGGTA GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
 223_7    CCAATCGGTA GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
  A3_4    CCAAGCGGTA GGACGTTCTT CATTCTACTG TCTAGAGTAT TTTCCCTCTC
  A3_5    CCAAGCGGTA GGACGTTCTT CATTCTACTG TCTAGAGTAT TTTCCCTCTC
  A3_7    CCAAGCGGTA GGACGTTCTT CATTCTACTG TCTAGAGTAT TTTCCCTCTC
  A3_3    CCAAGCGGTA GGACGTTCTT CATTCTACTG TCTAGAGTAT TTTCCCTCTC
 42_12    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
  AAV1    CCAAGCCGTG GGACGTTCAT CCTTTTACTG CCTGGAATAT TTCCCTTCTC
  AAV2    TCAGGCAGTA GGACGCTCTT CATTTTACTG CCTGGAGTAC TTTCCTTCTC
  AAV3    TCAAGCGGTG GGACGCTCAT CCTTTTACTG CCTGGAGTAC TTCCCTTCGC
  AAV8    TCAGGCCGTG GGACGCTCCT CCTTCTACTG CCTGGAATAC TTTCCTTCGC
  AAV9    TCAAGCGTTA GGACGTTCTT CTTTCTACTG TCTGGAATAC TTCCCTTCTC
  AAV7    TCAGTCTGTG GGACGTTCCT CCTTCTACTG CCTGGAGTAC TTCCCCTCTC
  44_2    TCAGGCCGTG GGCCGTTCCT CCTTCTACTG CCTGGAGTAC TTTCCTTCTC
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By: 
Attorneys

Fig. 1AAAA

	3901				3950
42_2	AGACAACGCT	GGAA.....	AACGTGCTAA	TGACCAGCGA	GGAGGAGATC
42_8	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_15	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_5b	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_1b	AGACAACG.T	AGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_13	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_3a	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_4	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_5a	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATC
42_10	AGACAACGCT	GGAA.....	AACGTGCTAA	TGACCAGCGA	GGAGGAGATC
42_3b	AGACAACGCT	GGAA.....	AACGTGCTAA	TGACCAGCGA	GGAGGAGATC
42_11	AGACAACGCT	GGAA.....	AACGTGCTAA	TGACCAGCGA	GGAGGAGATC
42_6b	AGACAACGCT	GGAA.....	AACGTGCTAA	TGACCAGCGA	GGAGGAGATC
43_1	AGACAATG.T	GGACTACAGC	AGCGTGATGC	TCACCAGCGA	AGAAGAAATT
43_5	AGACAATG.T	GGACTACAGC	AGCGTGATGC	TCACCAGCGA	AGAAGAAATT
43_12	AGACAATG.T	GGACTACAGC	AGCGTGATGC	TCACCAGCGA	AGAAGAAATT
43_20	CGATGGAG.T	GGATTACAGC	CAAGTGCTGA	TTACAGATGA	GGAAGAAATC
43_21	CGATGGAG.T	GGATTACAGC	CAAGTGCTGA	TTACAGATGA	GGAAGAAATC
43_23	CGATGGAG.T	GGATTACAGC	CAAGTGCTGA	TTACAGATGA	GGAAGAAATC
43_25	CGATGGAG.T	GGATTACAGC	CAAGTGCTGA	TTACAGATGA	GGAAGAAATC
44_1	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATT
44_5	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATT
223_10	AAACTACATT	AGAA.....	AACGTGCTCA	TGACAAATGA	AGAAGAAATT
223_2	AAACTACATT	AGAA.....	AACGTGCTCA	TGACAAATGA	AGAAGAAATT
223_4	AAACTACATT	AGAA.....	AACGTGCTCA	TGACAAATGA	AGAAGAAATT
223_5	AAACTACATT	AGAA.....	AACGTGCTCA	TGACAAATGA	AGAAGAAATT
223_6	AAACTACATT	AGAA.....	AACGTGCTCA	TGACAAATGA	AGAAGAAATT
223_7	AAACTACATT	AGAA.....	AACGTGCTCA	TGACAAATGA	AGAAGAAATT
A3_4	TACCAATG.T	GGACATTGAA	TCAGTGCTTA	TTACAGACGA	AGAAGAAATC
A3_5	TACCAATG.T	GGACATTGAA	TCAGTGCTTA	TTACAGACGA	AGAAGAAATC
A3_7	TACCAATG.T	GGACATTGAA	TCAGTGCTTA	TTACAGACGA	AGAAGAAATC
A3_3	TACCAATG.T	GGACATTGAA	TCAGTGCTTA	TTACAGACGA	AGAAGAAATC
42_12	AGACAACGCT	GGAA.....	AACGTGCTAA	TGACCAGCGA	GGAGGAGATC
AAV1	TTCAAACA.C	TGCATTGGAC	AATGTCATGA	TTACAGACGA	AGAGGAAATT
AAV2	AACAAATG.T	GAACATTGAA	AAGGTCATGA	TTACAGACGA	AGAGGAAATC
AAV3	AAGTAACG.C	AGAATTAGAT	AATGTAATGA	TTACGGATGA	AGAAGAGATT
AAV8	AGACAATG.C	GGATTACAGC	GATGTCATGC	TCACCAGCGA	GGAAGAAATC
AAV9	CGATGGAG.T	CGACTACAGC	CAGGTGCTGA	TTACAGATGA	GGAAGAAATT
AAV7	AAACTACATT	GGAA.....	AATGTGTTAA	TGACAAATGA	AGAAGAAATT
44_2	AGACAACG.T	GGACTATAGC	AGCGTTATGC	TAACCAGTGA	GGAAGAAATT

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 1AAAK

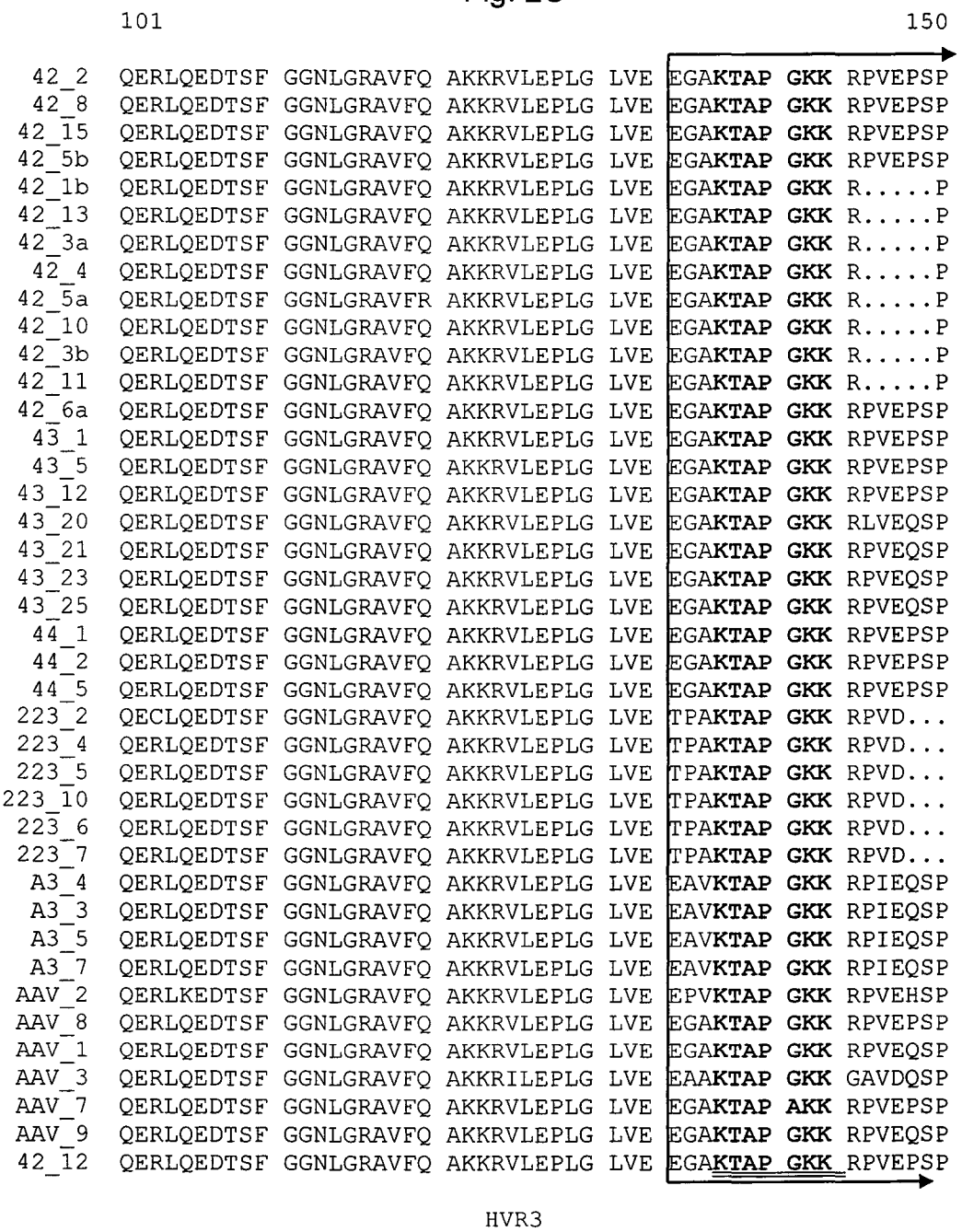
```
         4401                                                    4450
  42_2   AACAACGAAG GGGTTTATAC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
  42_8   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_15   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_5b   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_1b   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_13   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_3a   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
  42_4   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_5a   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_10   AACAACGAAG GGGTTTATAC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_3b   AACAACGAAG GGGTTTATAC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_11   AACAACGAAG GGGTTTATAC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
 42_6b   AACAACGAAG GGGTTTATAC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
  43_1   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CTCGTTATCT
  43_5   AATACCGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CTCGTTATCT
 43_12   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CTCGTTATCT
 43_20   AACACGGAAG GAGTTTATAG CGAGCCTCGC CCCATTGGCA CCCGTTACCT
 43_21   AACACGGAAG GAGTTTATAG CGAGCCTCGC CCCATTGGCA CCCGTTACCT
 43_23   AACACGGAAG GAGTTTATAG CGAGCCTCGC CCCATTGGCA CCCGTTACCT
 43_25   AACACGGAGG GGGTTTATAG CGAGCCTCGC CCCATTGGCA CCCGTTACCT
  44_1   AACACAGATG GCACTTATTC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
  44_5   AACACAGATG GCACTTATTC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
223_10   GACAGCCAGG GTGTTTACTC TGAGCCT... .......... ..........
 223_2   GACAGCCAGG GTGTTTACTC TGAGCCT... .......... ..........
 223_4   GACAGCCAGG GTGTTTACTC TGAGCCT... .......... ..........
 223_5   GACAGCCAGG GTGTTTACTC TGAGCCT... .......... ..........
 223_6   GACAGCCAGG GTGTTTACTC TGAGCCT... .......... ..........
 223_7   GACAGCCAGG GTGTTTACTC TGAGCCT... .......... ..........
  A3_4   GACGCAAACG GTGTTTATTC TGAACCCCGC CCTATTGGCA CTCGTTACCT
  A3_5   GACGCAAACG GTGTTTATTC TGAACCCCGC CCTATTGGCA CTCGTTACCT
  A3_7   GACGCAAACG GTGTTTATTC TGAACCCCGC CCTATTGGCA CTCGTTACCT
  A3_3   GACGCAAACG GTGTTTATTC TGAACCCCGC CCTATTGGCA CTCGTTACCT
 42_12   AATACTGAGG GTACTTATTC AGAGCCTCGC CCCATTGGCA CCCGTTACCT
   AAV1  GACAACAATG GACTTTATAC TGAGCCTCGC CCCATTGGCA CCCGTTACCT
   AAV2  GATACTAATG GCGTGTATTC AGAGCCTCGC CCCATTGGCA CCAGATACCT
   AAV3  GACACTAATG GTGTTTATAG TGAACCTCGC CCTATTGGAA CCCGGTATCT
   AAV8  AATACAGAAG GCGTGTACTC TGAACCCCGC CCCATTGGCA CCCGTTACCT
   AAV9  AATACCGAAG GTGTTTACTC TGAGCCTCGC CCCATTGGTA CTCGTTACCT
   AAV7  GACAGCCAGG GTGTTTACTC TGAGCCTCGC CCTATTGGCA CTCGTTACCT
  44_2   AACACAGATG GCACTTATTC TGAGCCTCGC CCCATCGGCA CCCGTTACCT
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 2C



THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys

Fig. 2M

```
      601                                                        650
 42_2   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 42_8   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_15   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_5b   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_1b   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_13   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_3a   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 42_4   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_5a   SQGALPGMAW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_10   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_3b   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_11   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_6a   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMDG FGLKHPPPQI
 43_1   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 43_5   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
43_12   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
43_20   NQGVIPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
43_21   NQGVIPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
43_23   NQGVIPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
43_25   NQGVIPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 44_1   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 44_2   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 44_5   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
223_2   NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
223_4   NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
223_5   NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
223_10  NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
223_6   NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
223_7   NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
 A3_4   SQGILPGMVW QDRDVYLQGP IWAKTPHTDG HFHPSPLMGG FGLKHPPPQI
 A3_3   SQGILPGMVW QDRDVYLQGP IWAKTPHTDG HFHPSPLMGG FGLKHPPPQI
 A3_5   SQGILPGMVW QDRDVYLQGP IWAKTPHTDG HFHPSPLMGG FGLKHPPPQI
 A3_7   SQGILPGMVW QDRDVYLQGP IWAKTPHTDG HFHPSPLMGG FGLKHPPPQI
AAV_2   TQGVLPGMVW QDRDVYLQGP IWAKIPHTDG HFHPSPLMGG FGLKHPPPQI
AAV_8   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
AAV_1   AMGALPGMVW QDRDVYLQGP IWAKIPHTDG HFHPSPLMGG FGLKNPPPQI
AAV_3   HQGALPGMVW QDRDVYLQGP IWAKIPHTDG HFHPSPLMGG FGLKHPPPQI
AAV_7   NQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
AAV_9   NQGVIPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
42_12   SQGALPGMVW QNRDVYLQGP IWAKIPHTDG NFHPSPLMGG FGLKHPPPQI
HVR10 (cont'd)
```

THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA
Applicant
SAPALO VELEZ BUNDANG & BULILAN

By:

Attorneys