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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 October 2003 (23.10.2003)

PCT

(10) International Publication Number
WO 03/087334 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US03/11191

(22) International Filing Date: 9 April 2003 (09.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/371,044 9 April 2002 (09.04.2002) US

(71) Applicant *(for all designated States except US)*: **CORNELL RESEARCH FOUNDATION, INC.** [US/US]; Cornell Business & Technology Park, 20 Thornwood Drive, Suite 105, Ithaca, NY 14850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FALCK-PEDERSEN, Eric, S.** [US/US]; 8714 Buena Vista Drive, Dobbs Ferry, NY 10522 (US). **PHILPOTT, Nicola** [GB/US]; 506 East 48th Street, Apt. 2W, New York, NY 10028 (US).

(74) Agents: **KILYK, John, Jr.** et al.; Leydig, Voit & Mayer, Ltd., Two Prudential Plaza, Suite 4900, 180 North Stetson, Chicago, IL 60601-6780 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/087334 A2

(54) Title: USE OF AAV INTEGRATION EFFICIENCY ELEMENT FOR MEDIATING SITE-SPECIFIC INTEGRATION OF A TRANSCRIPTION UNIT

CGACATTTTGCGACACCATGTGGTCACGCTGGGTATTTAAGCCCGAGTGAGCACGCAGGG 60
TCTCCATTTTGAAGCGGGAGGTTTGAACGCGCAGCCGCCATG 102
(SEQ ID NO:1)

(57) Abstract: The invention provides an expression construct comprising a nucleic acid sequence encoding an adeno-associated virus integration efficiency element (AAV IEE), wherein the expression construct is substantially devoid of AAV inverted terminal repeats (AAV ITRs). Such an expression construct site-specifically integrates into a host cell chromosome when provided to a host cell in conjunction with an AAV Rep protein. The invention also provides a method of integrating a nucleic acid sequence of interest into a host cell chromosome through use of such an expression construct, as well as a method of prophylactically or therapeutically treating a mammal for a pathologic state comprising administering to a mammal such an expression construct comprising a nucleic acid sequence encoding a therapeutic factor.

USE OF AAV INTEGRATION EFFICIENCY ELEMENT FOR MEDIATING SITE-SPECIFIC INTEGRATION OF A TRANSCRIPTION UNIT

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0001] This invention was made in part with United States Government support under Grant Number HL59312 awarded by the National Heart, Blood and Lung Institute of the National Institutes of Health. The United States Government may have certain rights in this invention.

TECHNICAL FIELD OF THE INVENTION

[0002] The invention pertains to an expression construct containing a nucleic acid sequence encoding an adeno-associated integration efficiency element to achieve site-specific integration of a nucleic acid sequence of interest into a host cell genome.

BACKGROUND OF THE INVENTION

[0003] Gene therapy is emerging as a popular form of treatment that aims to address a variety of disease states through the transfer of functional genetic material into cells. Critical to the success of gene therapy is the development of safe and efficient gene transfer vehicles. To date, various strategies have been developed for the transfer of therapeutic genes and include both viral and nonviral means. All of these gene delivery systems, however, suffer from limitations in their applicability and efficacy (see, e.g., Verma et al., *Nature,* 389:239-242 (1997)). The most commonly used gene transfer vehicles have been of viral origin (e.g., adenovirus, retrovirus, adeno-associated virus).

[0004] Adenovirus (Ad) vectors provide efficient means of transgene delivery to a variety of cell types, regardless of mitotic state. They do, however, produce only transient expression (no integration and no replication if replication-deficient) *in vivo* and may cause adverse reactions when administered. Retroviruses offer the desirable feature of being able to insert a gene of interest into the host genome, thus contributing to the stability of the transduced gene. However, retroviruses have a limited host range, and successful infection occurs only in mitotic cells, with the exception of the human immunodeficiency virus. Additionally, retroviruses integrate randomly into the host cell chromosome, thus raising some concern about the potential activation of transcriptionally silent oncogenes and the possible inactivation of tumor suppressor genes mediated by insertional mutagenesis. Adeno-associated viral (AAV) vectors, however, have not been associated with any human disease and are capable of site-specific integration into the host genome. The major disadvantages associated with AAV vectors are their inability to carry large transgene

sequences and the relatively low yield of recombinant vector produced by current methods. In view of these limitations, there have been many attempts to combine the site-specific integration properties of AAV vectors with other vector systems, such as Ad vectors (see, e.g., U.S. Patent 5,856,152) and baculovirus (see, e.g., Palombo et al., *J. Virol.,* 72(6):5025-5034 (1998), as well as with nonviral delivery systems, such as liposomes (see, e.g., Lamartina et al., *J. Virol.,* 72(9):7653-7658 (1998)).

[0005] Adeno-associated virus is a human parvovirus with a single-stranded DNA genome of 4.7 kb (see, e.g., Berns et al., *Bioessays,* 17:237-245 (1995)). It contains two open reading frames (ORFs), Rep and Cap, which are flanked by two inverted terminal repeats (ITRs) (see, e.g., Kotin, *Hum. Gene Ther.,* 5:793-801 (1994) and Srivastava, *Intervirology,* 27:138-147 (1987)). The ITRs are 160 nucleotides in length and are considered to be the only *cis* elements required for replication and site-specific integration of the AAV genome. In addition to the ITR elements, the Rep gene is necessary *in trans* to target the integration event to the AAVS1 site located on human chromosome 19 (see, e.g., Balague et al., *J. Virol.,* 71:3299-3306 (1997), Bertran et al., *Ann. NY Acad. Sci.,* 850:163-177 (1998), Lamartina et al. (1998), *supra*, Pieroni et al., *Virology,* 249:249-259 (1998), Shelling et al., *Gene Ther.,* 1:165-169 (1994), and Surosky et al., *J. Virol.,* 71:7951-7959 (1997)).

[0006] The Rep ORF encodes four non-structural proteins, Rep 40, 52, 68, and 78, which are involved in the replication of the AAV genome (see, e.g., Srivastava et al., *J. Virol.,* 45:555-564 (1983), and Trempe et al., *Virology,* 161:18-28 (1987)). Rep 68 and 78 are gene products of alternatively spliced mRNA transcribed from the AAV p5 promoter. These larger Rep proteins have DNA binding, site-specific and strand-specific endonuclease activities, as well as ATPase and DNA-DNA and DNA-RNA helicase functions (see, e.g., Im et al., *Cell,* 61:447-457 (1990), Im et al., *J. Virol.,* 66:1119-1128 (1992), Wonderling et al., *J. Virol.,* 70:4783-4786 (1996), and Zhou et al., *J. Virol.,* 73:1580-1590 (1999)). Transcription from the p19 promoter generates Rep 52 and Rep 40. Rep 52 is known to have helicase and ATPase activities but no DNA binding or endonuclease activity (see, e.g., Smith et al., *J. Virol.,* 72:4874-4881 (1998)).

[0007] As mentioned previously, AAV is the only known virus that site-specifically integrates into the human genome. Such site-specific integration occurs at the AAVS1 site at position 19q13.3 located on human chromosome 19 (see, e.g., Giraud et al., *PNAS,* 91:10039-10043 (1994), Kotin et al., *PNAS,* 87:2211-2215 (1990), and Samulski et al., *EMBO J.,* 10:3941-3950 (1991)). Although the precise molecular mechanisms of AAV site-specific integration are not well understood, it appears that Rep 68 and 78 are critical. Indeed, it is thought that these proteins function by binding to the Rep Binding Elements (RBEs) situated in both AAV ITRs and at the AAVS1 site (see, e.g., Chiorini et al., *Hum.*

Gene Ther., 6:1531-1541 (1995), Giraud et al. (1994), *supra*, and Weitzman et al., *J. Virol.*, PNAS, 91:5808-5812 (1994)). The endonuclease activity of the two larger Rep proteins allows them to nick at the terminal resolution site (trs), which is positioned 8 bp or 11 bp away from the RBEs in AAV and AAVS1, respectively. An interaction between Rep molecules that are bound to the AAV genome with those that are bound to the AAVS1 site then takes place, and a nonhomologous recombination event occurs, resulting in integration of the AAV genome (see, e.g., Linden et al., *PNAS*, 93:11288-11294 (1996), Linden et al., *PNAS*, 93:7966-7972 (1996), Urcelay et al., *J. Virol.*, 69:2038-2046 (1995), and Yang et al., *J. Virol.*, 71:9231-9247 (1997)). It has been shown that head-to-tail concatemers of the wild-type AAV genome are able to site-specifically integrate in this manner (see, e.g., Giraud et al., *J. Virol.*, 69:6917-6924 (1995)).

[0008] There have been numerous attempts to modify current vector systems to include the site-specific integration properties of AAV. A majority of these strategies have involved incorporating an expression cassette comprising a transgene flanked by at least one AAV ITR into a replication deficient Ad vector. When administered to a mammal in conjunction with an AAV Rep protein provided *in trans*, the expression cassette integrates into the mammalian genome. Other attempts have involved creating recombinant AAV (rAAV) vectors by modifying (e.g., mutating) one or more of the AAV ITRs. For example, International Patent Application WO 99/64569 describes methods and compositions for generating rAAV vectors. The methods generally comprise providing a helper plasmid encoding Rep and Cap polypeptides, providing a rAAV plasmid, which generally comprises a heterologous nucleotide sequence flanked by two AAV ITRs, and introducing both into a host cell. The distinction of this method, as compared to the earlier methods, purportedly is that there is no distal D-sequence homology between the helper plasmid and the rAAV plasmid. The D-sequence is contained in the AAV ITRs and is comprised of a stretch of 20 nucleotides. The '569 PCT application contends that the distal 10 nucleotides of the D-sequence are responsible for wild-type AAV contamination and that the proximal 10 nucleotides are necessary and sufficient to mediate high-efficiency rescue, replication, and encapsidation of the viral genome *in vivo*. In that respect, the rAAV plasmid may be designed to lack the distal 10 nucleotides of the D-sequence.

[0009] Other experiments involving AAV vectors with mutant ITRs were performed by Young et al., *PNAS*, 98(24):13525-13530 (2001). These mutants contain a nonfunctional terminal resolution site (trs), which renders the AAV vectors incapable of supporting viral replication. The results presented by Young et al. demonstrate that, in the presence of Rep 78, both wild-type AAV and the trs mutant AAV vectors target to chromosome 19 with similar frequency. Young et al. thus suggest that a functional ITR, as defined by AAV lytic replication, is not required for targeted integration. In that respect, these AAV vectors can

be designed to contain mutant ITRs that are defective for AAV Rep-dependent replication but are completely competent for site-specific integration.

[0010] While previously described vectors have been somewhat effective at achieving site-specific integration, a need remains to provide an expression construct which is more efficient at integrating into a specific location into a host cell chromosome, such that these expression constructs can be formulated into therapeutic compositions and used in a method of treating a variety of pathologic states. The invention provides such a construct, composition, and method. These and other advantages of the invention, as well as additional inventive features, will be apparent from the description of the invention provided herein.

SUMMARY OF THE INVENTION

[0011] The invention provides an expression construct comprising a nucleic acid sequence encoding an adeno-associated virus integration efficiency element (AAV IEE), wherein the expression construct is substantially devoid of AAV inverted terminal repeats (AAV ITRs). Such an expression construct site-specifically integrates into a host cell chromosome when provided to a host cell in conjunction with an AAV Rep protein. The invention also provides a method of integrating a nucleic acid sequence of interest into a host cell chromosome. The method comprises providing to a host cell an expression construct comprising a nucleic acid sequence encoding an AAV IEE and a nucleic acid sequence of interest, wherein the expression construct is substantially devoid of AAV ITRs, and providing to the host cell a nucleic acid sequence encoding an AAV Rep protein. The nucleic acid sequence encoding the AAV Rep protein is expressed, thereby resulting in the site-specific integration of the nucleic acid sequence of interest into a chromosome contained in the host cell. Further provided by the invention is a method of prophylactically or therapeutically treating a mammal for a pathologic state. The method comprises administering to a mammal an expression construct comprising a nucleic acid sequence encoding an AAV IEE and a nucleic acid sequence encoding a therapeutic factor, wherein the expression construct is substantially devoid of AAV ITRs, and administering to the mammal a nucleic acid sequence encoding an AAV Rep protein. The nucleic acid sequence encoding the AAV Rep protein is expressed, thereby resulting in the site-specific integration of the nucleic acid sequence encoding a therapeutic factor into a chromosome of a cell of the mammal, the expression of the nucleic acid sequence encoding the therapeutic factor, and the subsequent production of the therapeutic factor to prophylactically or therapeutically treat the mammal for the pathologic state.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Fig. 1 sets forth the nucleotide sequence encoding an adeno-associated integration efficiency element isolated from the adeno-associated virus of serotype 2 (SEQ ID NO:1).

[0013] Fig. 2 sets forth the homology of AAV IEE isolated from the adeno-associated virus of serotype 2 with regions of the adeno-associated virus of serotypes 1 (SEQ ID NO:3), 3 (SEQ ID NO:4), 4 (SEQ ID NO:5), 5 (SEQ ID NO:6), and 6 (SEQ ID NO:2)).

DETAILED DESCRIPTION OF THE INVENTION

[0014] The invention is predicated on the discovery of a nucleic acid sequence element which, when provided in collaboration with the necessary Rep proteins, is sufficient for Rep-mediated site-specific integration into a host cell chromosome. The invention is contrary to the currently held belief that site-specific integration is possible only in the presence of at least one adeno-associated virus inverted terminal repeat (AAV ITR). Accordingly, the invention provides an expression construct comprising a nucleic acid sequence encoding an adeno-associated virus integration efficiency element (AAV IEE), wherein the expression construct is substantially devoid of AAV ITRs, and wherein the expression construct site-specifically integrates into a host cell chromosome when provided to the host cell in conjunction with an AAV Rep protein.

[0015] By "substantially devoid of AAV ITRs" is meant lacking at least the portion of an AAV ITR that is responsible for integration. AAV ITRs also are responsible for a number of other functions, including serving as the origins of replication for viral DNA synthesis. Thus, it is conceivable to include these regions in the expression construct but to exclude the regions(s) responsible for integration. Preferably, however, the AAV ITRs are completely removed from the expression constructs of the invention. The AAV IEE sequence can be included in an expression construct containing ITRs other than AAV ITRs (e.g., adenoviral ITRs).

[0016] AAV IEE is part of the p5 promoter region of AAV and is comprised of a nucleotide sequence containing various transcription factor-binding sites. In this respect, it is believed that AAV IEE uses cellular transcription factors (e.g., YY1) in collaboration with Rep 68 and/or Rep 78 to form an active integration complex. It is also believed that AAV IEE functions as a promoter for RNA polymerase II transcription of Rep 68 and 78 transcripts.

[0017] The nucleic acid sequence encoding an AAV IEE of the invention can be isolated from any AAV of the *Dependovirus* genus. For example, an AAV IEE can comprise nucleotides 222-324 in wild-type (wt) AAV of serotype 2 (AAV 2) (SEQ ID NO:1) (**Fig.1**). This region, in particular, has been shown to be sufficient for directing site-specific integration in the absence of AAV ITRs. This nucleotide sequence also has shown to be

homologous to regions contained in other AAV serotypes as well (e.g., serotypes 1 (SEQ ID NO:3), 3 (SEQ ID NO:4), 4 (SEQ ID NO:5), 5 (SEQ ID NO:6), and 6 (SEQ ID NO:2)) (**Fig. 2**), thus suggesting an AAV IEE exists in these other serotypes. Accordingly, a nucleic acid sequence encoding an AAV IEE, as it is referred to herein, is meant to encompass nucleotides 222-324 of wt AAV 2, homologous regions to this AAV 2 region in any other AAV serotype, such as AAV serotypes 1, 3, 4, 5, and 6, as well as fragments of any of the foregoing that direct site-specific integration. Performing routine experimentation can identify homologous regions and fragments suitable for use in the invention. Such experimentation will typically involve isolating nucleotides 222-324 of wt AAV 2, identifying homologous regions in other AAV serotypes, generating fragments of an AAV IEE from any AAV serotype, and assaying for activity of the homologous regions and fragments (e.g., assaying for the involvement of a particular homologous region or fragment in site-specific integration).

[0018] The term "host cell" denotes any cell that can be, or has been, used as a recipient of a nucleic acid sequence or an expression construct, and includes the progeny of the original cell which has been infected with the expression construct. These host cells must include a chromosome that allows site-specific integration of an expression construct of the invention. Thus, a host cell of the invention generally refers to a mammalian cell (e.g., a human cell). It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

[0019] In the context of the invention, in terms of provision of the various aspects of the invention to cells, expression constructs are provided to a host cell, which is preferably a eukaryotic host cell. The eukaryotic host cell can be present *in vitro* or *in vivo*. According to the invention, "contacting" of cells with an expression construct of the invention can be by any suitable means by which the expression construct will be introduced into the cell. Preferably the expression constructs will be introduced by infection using the natural capability of the expression construct to enter cells (e.g., the capability of an expression construct to enter cells via receptor-mediated endocytosis). However, the expression constructs can be introduced by any other suitable means, e.g., transfection, calcium phosphate-mediated transformation, microinjection, electroporation, osmotic shock, and the like. Similarly, in a preferred embodiment of the invention, *in vivo* transfer of expression constructs is contemplated. Accordingly, the method of the invention also contemplates transfer *in vivo* by the methods set forth herein, or by any suitable method.

[0020] The AAV Rep protein used in the context of the invention can be any AAV Rep protein or combination of AAV Rep proteins, which is capable of providing the necessary function(s) to allow for site specific integration. Typically, AAV Rep proteins are

characterized as being a short or a long form of Rep. The term "long forms of Rep" refers to the Rep 78 and/or Rep 68 gene products of the AAV Rep coding region, including functional homologues thereof. The long forms of Rep are normally expressed under the direction of the AAV p5 promoter. The term "short forms of Rep" refers to the Rep 52 and/or Rep 40 gene products of the AAV Rep coding region, including functional homologues thereof. The short forms of Rep are expressed under the direction of the AAV p19 promoter. Preferably, the AAV Rep protein comprises a long form of Rep (i.e., Rep 68 and/or Rep 78).

[0021] The AAV Rep protein can be provided to the host cell by any suitable means. For example, the AAV Rep protein can be provided to the host cell *in trans*. By "*in trans*" is meant being provided to the host cell via a different DNA molecule than the DNA molecule comprising the expression construct. However, it is also suitable for the AAV Rep protein to be provided to the host cell in the expression construct itself or, alternatively, incorporated into a packaging cell line. AAV Cap proteins also may be provided to the host cell by any suitable means, if desired. A number of vectors that contain the AAV Rep coding region are known, including those vectors described in U.S. Patent 5,139,941, having ATCC accession numbers 53222, 53223, 53224, 53225 and 53226. Similarly, methods of obtaining vectors containing the HHV-6 homologue of AAV Rep are described in Thomson et al., *Virology,* 204:304-311 (1994). A number of vectors containing the AAV Cap coding region have also been described, including those vectors described in U.S. Patent 5,139,941. Packaging cell lines derived from human 293 cells that have been transfected with a vector having the AAV Rep gene operably linked to a heterologous transcription promoter have been described in International Publication Nos. WO 95/13392 and WO 95/13365.

[0022] One of ordinary skill in the art will appreciate that any of a number of expression constructs known in the art are suitable for use in the invention. Examples of suitable expression constructs include, for instance, plasmids, plasmid-liposome complexes, and viral vectors. Any of these expression constructs are capable of being manipulated to include a nucleic acid sequence encoding an AAV IEE and can be prepared using standard recombinant DNA techniques described in, e.g., Sambrook et al., Molecular Cloning, a Laboratory Manual, 2d edition, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989), and Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates and John Wiley & Sons, New York, N.Y. (1994). Alternatively, the nucleic acid sequence encoding an AAV IEE can be administered to a host cell as naked DNA.

[0023] Plasmids are genetically engineered circular double-stranded DNA molecules and can be designed to contain an expression cassette comprising an AAV IEE. Although plasmids were the first vector described for administration of therapeutic nucleic acids, the

level of transfection efficiency is poor compared with other techniques. By complexing the plasmid with liposomes, the efficiency of gene transfer in general is improved. While the liposomes used for plasmid-mediated gene transfer strategies have various compositions, they are typically synthetic cationic lipids. Advantages of plasmid-liposome complexes include their ability to transfer large nucleic acid sequences and their relatively low immunogenicity. While plasmids are suitable for use in the invention, preferably the expression construct is a viral vector.

[0024] The viral vector can be any suitable viral vector. Suitable viral vectors include, but are not limited to, reoviruses, adenoviruses, adeno-associated based viruses, papovaviruses, parvoviruses, picornaviruses, and enteroviruses of any suitable origin (preferably of animal origin (e.g., avian or mammalian) and desirably of human origin). By "adeno-associated based viruses" is meant an expression construct containing any AAV derived gene(s) excluding the AAV ITRs. Other suitable viral vectors are known in the art and are well characterized. Examples of such viral vectors are described in, for example, Fields et al., VIROLOGY Lippincott-Raven (3rd ed. (1996) and 4th ed. (2000)); ENCYCLOPEDIA OF VIROLOGY, R.G. Webster et al., eds., Academic Press (2nd ed., 1999); FUNDAMENTAL VIROLOGY, Fields et al., eds., Lippincott-Raven (3rd ed., 1995); Levine, "Viruses," Scientific American Library No. 37 (1992); MEDICAL VIROLOGY, D.O. White et al., eds., Academic Press (2nd ed. 1994); and INTRODUCTION TO MODERN VIROLOGY, Dimock, N.J. et al., eds., Blackwell Scientific Publications, Ltd. (1994). Preferably, the viral vector is derived from, or based on, a virus that normally infects animals, such as mammals (most preferably humans). Adenoviral (Ad) vectors based on human adenoviruses are preferred viral vectors.

[0025] Adenovirus is a 36 kb double-stranded DNA virus that efficiently transfers DNA *in vivo* to a variety of different target cell types. The Ad vector can be produced in high titers and can efficiently transfer DNA to replicating and non-replicating cells. The Ad vector genome can be generated using any species, strain, subtype, mixture of species, strains, or subtypes, or chimeric adenovirus as the source of vector DNA. Adenoviral stocks that can be employed as a source of adenovirus can be amplified from the adenoviral serotypes 1 through 51, which are currently available from the American Type Culture Collection (ATCC, Manassas, VA), or from any other serotype of adenovirus available from any other source. For instance, an adenovirus can be of subgroup A (e.g., serotypes 12, 18, and 31), subgroup B (e.g., serotypes 3, 7, 11, 14, 16, 21, 34, and 35), subgroup C (e.g., serotypes 1, 2, 5, and 6), subgroup D (e.g., serotypes 8, 9, 10, 13, 15, 17, 19, 20, 22-30, 32, 33, 36-39, and 42-47), subgroup E (serotype 4), subgroup F (serotypes 40 and 41), or any other adenoviral serotype. Given that the human adenovirus serotype 5 (Ad5) genome has been completely sequenced, the adenoviral vector is described herein with

respect to the Ad5 serotype. The Ad vector can be any adenoviral vector capable of growth in a cell, which is in some significant part (although not necessarily substantially) derived from or based upon the genome of an adenovirus. The Ad vector can be based on the genome of any suitable wild-type adenovirus. Preferably, the Ad vector is derived from the genome of a wild-type adenovirus of group C, especially of serotype 2 or 5. Ad vectors are well known in the art and are described in, for example, U.S. Patents 5,559,099, 5,712,136, 5,731,190, 5,837,511, 5,846,782, 5,851,806, 5,962,311, 5,965,541, 5,981,225, 5,994,106, 6,020,191, and 6,113,913, International Patent Applications WO 95/34671, WO 97/21826, and WO 00/00628, and Thomas Shenk, "Adenoviridae and their Replication," and M. S. Horwitz, "Adenoviruses," Chapters 67 and 68, respectively, in *Virology,* B. N. Fields et al., eds., 3d ed., Raven Press, Ltd., New York (1996).

[0026] Preferably, the Ad vector is replication-deficient. By "replication-deficient" is meant that the Ad vector comprises a genome that lacks at least one replication-essential gene function. A deficiency in a gene, gene function, or gene or genomic region, as used herein, is defined as a deletion of sufficient genetic material of the viral genome to impair or obliterate the function of the gene whose nucleic acid sequence was deleted in whole or in part. Replication-essential gene functions are those gene functions that are required for replication (i.e., propagation) of a replication-deficient Ad vector. Replication-essential gene functions are encoded by, for example, the adenoviral early regions (e.g., the E1, E2, and E4 regions), late regions (e.g., the L1-L5 regions), genes involved in viral packaging (e.g., the IVa2 gene), and virus-associated RNAs (e.g., VA-RNA I and/or VA-RNA II). Preferably, the replication-deficient Ad vector comprises an adenoviral genome deficient in two or more gene functions required for viral replication. The two or more regions of the adenoviral genome are preferably selected from the group consisting of the E1, E2, and E4 regions. More preferably, the replication-deficient adenoviral vector comprises a deficiency in at least one replication-essential gene function of the E1 region (denoted an E1-deficient adenoviral vector). The E1 region of the adenoviral genome comprises the E1A region and the E1B region. The E1A and E1B regions comprise nucleic acid sequences coding for multiple peptides by virtue of RNA splicing. A deficiency of a gene function encoded by either or both of the E1A and/or E1B regions of the adenoviral genome (e.g., a peptide that performs a function required for replication) is considered a deficiency of a gene function of the E1 region in the context of the invention. In addition to such a deficiency in the E1 region, the recombinant adenovirus also can have a mutation in the major late promoter (MLP), as discussed in International Patent Application WO 00/00628. More preferably, the vector is deficient in at least one replication-essential gene function of the E1 region and at least part of the nonessential E3 region (e.g., an Xba I deletion of the E3 region) (denoted an E1/E3-deficient adenoviral vector).

[0027] Preferably, the adenoviral vector is "multiply deficient," meaning that the adenoviral vector is deficient in one or more gene functions required for viral replication in each of two or more regions of the adenoviral genome. For example, the aforementioned E1-deficient or E1/E3-deficient Ad vector can be further deficient in at least one replication-essential gene function of the E4 region (denoted an E1/E4-deficient adenoviral vector). An adenoviral vector deleted of the entire E4 region can elicit a lower host immune response.

[0028] Alternatively, the Ad vector lacks replication-essential gene functions in all or part of the E1 region and all or part of the E2 region (denoted an E1/E2-deficient adenoviral vector). Ad vectors lacking replication-essential gene functions in all or part of the E1 region, all or part of the E2 region, and all or part of the E3 region also are contemplated herein. If the Ad vector is deficient in a replication-essential gene function of the E2A region, the vector preferably does not comprise a complete deletion of the E2A region, which is less than about 230 base pairs in length. Generally, the E2A region of the adenovirus codes for a DBP (DNA binding protein), a polypeptide required for DNA replication. DBP is composed of 473 to 529 amino acids depending on the viral serotype. It is believed that DBP is an asymmetric protein that exists as a prolate ellipsoid consisting of a globular Ct with an extended Nt domain. Studies indicate that the Ct domain is responsible for DBP's ability to bind to nucleic acids, bind to zinc, and function in DNA synthesis at the level of DNA chain elongation. However, the Nt domain is believed to function in late gene expression at both transcriptional and post-transcriptional levels, is responsible for efficient nuclear localization of the protein, and also may be involved in enhancement of its own expression. Deletions in the Nt domain between amino acids 2 to 38 have indicated that this region is important for DBP function (Brough et al., *Virology*, *196*, 269-281 (1993)). While deletions in the E2A region coding for the Ct region of the DBP have no effect on viral replication, deletions in the E2A region which code for amino acids 2 to 38 of the Nt domain of the DBP impair viral replication. It is preferable that the multiply replication-deficient adenoviral vector contain this portion of the E2A region of the adenoviral genome. In particular, for example, the desired portion of the E2A region to be retained is that portion of the E2A region of the adenoviral genome which is defined by the 5' end of the E2A region, specifically positions Ad5(23816) to Ad5(24032) of the E2A region of the adenoviral genome of serotype Ad5.

[0029] The Ad vector can be deficient in replication-essential gene functions of only the early regions of the adenoviral genome, only the late regions of the adenoviral genome, and both the early and late regions of the adenoviral genome. The adenoviral vector also can have essentially the entire adenoviral genome removed, in which case it is preferred that at least either the viral (i.e., adenoviral) inverted terminal repeats (Ad ITRs) and one or more

promoters or the Ad ITRs and a packaging signal are left intact (i.e., an adenoviral amplicon). The larger the region of the adenoviral genome that is removed, the larger the piece of exogenous nucleic acid sequence that can be inserted into the genome. For example, given that the adenoviral genome is 36 kb, by leaving the Ad ITRs and one or more promoters intact, the exogenous insert capacity of the adenovirus is approximately 35 kb. Alternatively, a multiply deficient Ad vector that contains only an Ad ITR and a packaging signal effectively allows insertion of an exogenous nucleic acid sequence of approximately 37-38 kb. Of course, the inclusion of a spacer element in any or all of the deficient adenoviral regions will decrease the capacity of the adenoviral vector for large inserts. Suitable replication-deficient Ad vectors, including multiply deficient Ad vectors, are disclosed in U.S. Patents 5,851,806 and 5,994,106 and International Patent Applications WO 95/34671 and WO 97/21826. An especially preferred adenoviral vector for use in the invention is that described in International Patent Application PCT/US01/20536.

[0030] It should be appreciated that the deletion of different regions of the Ad vector can alter the immune response of the mammal. In particular, the deletion of different regions can reduce the inflammatory response generated by the Ad vector. Furthermore, the Ad vector's coat protein can be modified so as to decrease the Ad vector's ability or inability to be recognized by a neutralizing antibody directed against the wild-type coat protein, as described in International Patent Application WO 98/40509.

[0031] The adenoviral vector, when multiply replication-deficient, especially in replication-essential gene functions of the E1 and E4 regions, preferably includes a spacer element to provide viral growth in a complementing cell line similar to that achieved by singly replication deficient Ad vectors, particularly an Ad vector comprising a deficiency in the E1 region. A spacer sequence is defined in the invention as any sequence of sufficient length to restore the size of the adenoviral genome to approximately the size of a wild-type adenoviral genome, such that the Ad vector is efficiently packaged into viral particles. The spacer element can contain any sequence or sequences which are of the desired length. The spacer element sequence can be coding or non-coding and native or non-native with respect to the adenoviral genome, but does not restore the replication-essential function to the deficient region. The spacer can be of any suitable size, desirably at least about 15 base pairs (e.g., between about 15 base pairs and about 12,000 base pairs), preferably about 100 base pairs to about 10,000 base pairs, more preferably about 500 base pairs to about 8,000 base pairs, even more preferably about 1,500 base pairs to about 6,000 base pairs, and most preferably about 2,000 to about 3,000 base pairs. The size of the spacer is limited only by the size of the insert that the Ad vector will accommodate (e.g., approximately 38 base pairs). In the absence of a spacer, production of fiber protein and/or viral growth of the multiply replication-deficient Ad vector is reduced by comparison to that of a singly

replication-deficient Ad vector. However, inclusion of the spacer in at least one of the deficient adenoviral regions, preferably the E4 region, can counteract this decrease in fiber protein production and viral growth. The use of a spacer in an Ad vector is described in U.S. Patent 5,851,806.

[0032] The Ad vector preferably contains a packaging domain. The packaging domain can be located at any position in the adenoviral genome, so long as the adenoviral genome is packaged into adenoviral particles. Preferably, the packaging domain is located downstream of the E1 region. More preferably, the packaging domain is located downstream of the E4 region. In a particularly preferred embodiment, the replication-deficient Ad vector lacks all or part of the E1 region and the E4 region. In this preferred embodiment, a spacer is inserted into the E1 region, a desired exogenous nucleic acid sequence of interest (e.g., a nucleic acid sequence encoding TNF-α) is located in the E4 region, and the packaging domain is located downstream of the E4 region. Thus, by relocating the packaging domain, the amount of potential overlap between the Ad vector and the cellular/helper virus genome is reduced.

[0033] The coat proteins of the Ad vector can be manipulated to alter the binding specificity of the resulting adenoviral particle. Suitable modifications to the coat proteins include, but are not limited to, insertions, deletions, or replacements in the adenoviral fiber, penton, pIX, pIIIa, pVI, or hexon proteins, or any suitable combination thereof, including insertions of various native or non-native ligands into portions of such coat proteins. Examples of Ad vectors with modified binding specificity are described in, e.g., U.S. Patents 5,871,727, 5,885,808, and 5,922,315. Preferred modified Ad vector particles include those described in, for example, Wickham et al., *J. Virol., 71(10),* 7663-9 (1997), Cripe et al., *Cancer Res., 61(7),* 2953-60 (2001), van Deutekom et al., *J. Gene Med., 1(6),* 393-9 (1999), McDonald et al., *J. Gene Med., 1(2),* 103-10 (1999), Staba et al., *Cancer Gene Ther., 7(1),* 13-9 (2000), Wickham, *Gene Ther., 7(2),* 110-4 (2000), Kibbe et al., *Arch. Surg., 135(2),* 191-7 (2000), Harari et al., *Gene Ther., 6(5),* 801-7 (2000), Bouri et al., *Hum Gene Ther., 10(10),* 1633-40 (1999), Wickham et al., *Nat. Biotechnol., 14(11),* 1570-3 (1996), Wickham et al., *Cancer Immunol. Immunother., 45(3-4),* 149-51 (1997), and Wickham et al., *Gene Ther., 2(10),* 750-6 (1995), and U.S. Patents 5,559,099; 5,712,136; 5,731,190; 5,770,442; 5,801,030; 5,846,782; 5,962,311; 5,965,541; 6,057,155; 6,127,525; and 6,153,435; and International Patent Applications WO 96/07734, WO 96/26281, WO 97/20051, WO 98/07865, WO 98/07877, WO 98/40509, WO 98/54346, WO 00/15823, and WO 01/58940.

[0034] Replication-deficient Ad vectors are typically produced in complementing cell lines that provide gene functions not present in the replication-deficient Ad vectors, but required for viral propagation, at appropriate levels in order to generate high titers of viral

vector stock. A preferred cell line complements for at least one and preferably all replication-essential gene functions not present in a replication-deficient adenovirus. The complementing cell line can complement for a deficiency in at least one replication-essential gene function encoded by the early regions, late regions, viral packaging regions, virus-associated RNA regions, or combinations thereof, including all adenoviral functions (e.g., to enable propagation of adenoviral amplicons, which comprise minimal adenoviral sequences, such as only Ad ITRs and the packaging signal or only Ad ITRs and an adenoviral promoter). Most preferably, the complementing cell line complements for a deficiency in at least one replication-essential gene function (e.g., two or more replication-essential gene functions) of the E1 region of the adenoviral genome, particularly a deficiency in a replication-essential gene function of each of the E1A and E1B regions. In addition, the complementing cell line can complement for a deficiency in at least one replication-essential gene function of the E2 (particularly as concerns the adenoviral DNA polymerase and terminal protein) and/or E4 regions of the adenoviral genome. Desirably, a cell that complements for a deficiency in the E4 region comprises the E4-ORF6 gene sequence and produces the E4-ORF6 protein. Such a cell desirably comprises at least ORF6 and no other ORF of the E4 region of the adenoviral genome. The cell line preferably is further characterized in that it contains the complementing genes in a non-overlapping fashion with the adenoviral vector, which minimizes, and practically eliminates, the possibility of the vector genome recombining with the cellular DNA. Accordingly, the presence of replication competent adenoviruses (RCA) is minimized if not avoided in the vector stock, which, therefore, is suitable for certain therapeutic purposes, especially gene therapy purposes. The lack of RCA in the vector stock avoids the replication of the Ad vector in non-complementing cells. The construction of complementing cell lines involves standard molecular biology and cell culture techniques, such as those described by Sambrook et al. (1989), *supra*, and Ausubel et al. (1984), *supra*. Complementing cell lines for producing the adenoviral vector include, but are not limited to, 293 cells (described in, e.g., Graham et al., *J. Gen. Virol.*, *36*, 59-72 (1977)), PER.C6 cells (described in, e.g., International Patent Application WO 97/00326, and U.S. Patents 5,994,128 and 6,033,908), and 293-ORF6 cells (described in, e.g., International Patent Application WO 95/34671 and Brough et al., *J. Virol.*, *71*, 9206-9213 (1997)).

[0035] The selection of expression construct for use in the invention will depend on a variety of factors such as, for example, the host, immunogenicity of the expression construct, the desired duration of protein production, the target cell, and the like. As each type of expression construct has distinct properties, a researcher has the freedom to tailor the invention to any particular situation. Moreover, more than one type of expression construct can be used, if desired.

[0036] The expression construct of the invention can further comprise a nucleic acid sequence of interest. Preferably, the nucleic acid sequence of interest is an exogenous nucleic acid sequence and encodes a protein. The nucleic acid sequence of interest can encode any protein that is desired for site-specific integration into a host cell genome. Preferably, the protein is a therapeutic factor useful to prophylactically or therapeutically treat a mammal for a pathologic state. For example, the therapeutic factor can be a cytokine. Examples of cytokines include, but are not limited to, interleukins, interferons (i.e., INF-α, INF-b, INF-γ), leukemia inhibitory factor (LIF), oncostatin M (OSM), granulocyte-macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), tumor necrosis factor-alpha (TNF- α), tumor necrosis factor-beta (TNF-β), and transforming growth factor-beta (TGF-β). Preferably, the cytokine is selected from the group consisting of an interleukin, an interferon and a tumor necrosis factor. The therapeutic factor also can be a protein that is toxic to a specific subset of cells. In this respect, the nucleic acid sequence can encode an apoptotic factor (e.g., Bax, Bak, Bcl-X_s, Bad, Bim, Bik, Bid, Harakiri, ICE-CED3 proteases, TRAIL, SARP-2, apoptin); an enzyme (e.g., cytosine deaminase, adenosine deaminase, hypoxanthine-guanine phosphoribosyltransferase, and thymidine kinase); a toxin (e.g., ricin A-chain, diphtheria toxin A, pertussis toxin A subunit, *E. coli* enterotoxin A subunit, cholera toxin A subunit and pseudomonas toxin c-terminal); an antisense molecule; a ribozyme; or a cell cycle regulator (e.g., p27, p21, p57, p18, p73, p19, p15, E2F-1, E2F-2, E2F-3, p107, p130 and E2F-4). Other therapeutic factors include those involved in the promotion or inhibition of angiogenesis. Alternatively, the nucleic acid sequence of interest can be useful for creating a stable host cell line wherein the nucleic acid sequence is permanently integrated into a specific location of the host cell genome.

[0037] In view of the above, the invention also provides a method of integrating a nucleic acid sequence of interest into a host cell chromosome. The method comprises (a) providing to a host cell an expression construct comprising a nucleic acid sequence encoding an AAV IEE and a nucleic acid sequence of interest, wherein the expression construct is substantially devoid of AAV ITRs, and (b) providing to the host cell a nucleic acid sequence encoding an AAV Rep protein, such that the nucleic acid sequence encoding the AAV Rep protein is expressed, thereby resulting in the site-specific integration of the nucleic acid sequence of interest into a chromosome contained in the host cell. Such a method can be useful for creating stable cell lines as well as for therapeutic purposes. For example, a host cell can be contacted with an expression construct as described above and then propagated to produce a stable host cell line. Such a cell line will have the nucleic acid sequence of interest permanently integrated into its genome in a specific location. Depending on the host cell, various propagation techniques can be employed. Host cell

propagation is well within the skill in the art and is described in, for example, Sambrook et al. (1989), *supra*.

[0038] The nucleic acid sequence of interest can be located at any suitable position in the expression construct. For example, the nucleic acid sequence can be positioned upstream of an AAV IEE. Preferably, however, the nucleic acid sequence of interest is positioned downstream of an AAV IEE in the expression construct. For the purpose of describing the relative position of nucleotide sequences in an expression construct throughout the instant application, such as when a particular nucleic acid sequence is described as being situated "upstream," "downstream," "3'," or "5'" relative to another sequence, it is to be understood that it is the position of the sequences in the "sense" or "coding" strand of a DNA molecule that is being referred to as is conventional in the art.

[0039] According to the invention, the nucleic acid sequence of interest is operably linked to regulatory sequences necessary for expression, i.e., a promoter. A "promoter" is a DNA sequence that directs the binding of RNA polymerase and thereby promotes RNA synthesis. A nucleic acid sequence is "operably linked" to a promoter when the promoter is capable of directing transcription of that nucleic acid sequence. A promoter can be native or non-native to the nucleic acid sequence to which it is operably linked.

[0040] Any promoter (i.e., whether isolated from nature or produced by recombinant DNA or synthetic techniques) can be used in connection with the invention to provide for transcription of a particular nucleic acid sequence. The promoter preferably is capable of directing transcription in a eukaryotic (desirably mammalian) cell. The functioning of the promoter can be altered by the presence of one or more enhancers and/or silencers present on the vector. "Enhancers" are *cis*-acting elements of DNA that stimulate or inhibit transcription of adjacent genes. An enhancer that inhibits transcription also is termed a "silencer." Enhancers differ from DNA-binding sites for sequence-specific DNA binding proteins found only in the promoter (which also are termed "promoter elements") in that enhancers can function in either orientation, and over distances of up to several kilobase pairs (kb), even from a position downstream of a transcribed region.

[0041] The invention preferentially employs a viral promoter. Suitable viral promoters are known in the art and include, for instance, cytomegalovirus (CMV) promoters, such as the CMV immediate-early promoter, promoters derived from human immunodeficiency virus (HIV), such as the HIV long terminal repeat promoter, Rous sarcoma virus (RSV) promoters, such as the RSV long terminal repeat, mouse mammary tumor virus (MMTV) promoters, HSV promoters, such as the Lap2 promoter or the herpes thymidine kinase promoter (Wagner et al., *PNAS*, 78, 144-145 (1981)), promoters derived from SV40 or Epstein Barr virus, an adeno-associated viral promoter, such as the p5 promoter, and the

like. Preferably, the viral promoter is an adenoviral promoter, such as the Ad2 or Ad5 major late promoter and tripartite leader, a CMV promoter, or an RSV promoter.

[0042] Many of the above-described promoters are constitutive promoters. Instead of being a constitutive promoter, the promoter can be an inducible promoter, i.e., a promoter that is up- and/or down-regulated in response to appropriate signals. Examples of suitable inducible promoter systems include, but are not limited to, the IL-8 promoter, the metallothionine inducible promoter system, the bacterial lacZYA expression system, the tetracycline expression system, and the T7 polymerase system. Further, promoters that are selectively activated at different developmental stages (e.g., globin genes are differentially transcribed from globin-associated promoters in embryos and adults) can be employed. The promoter sequence that regulates expression of the nucleic acid sequence can contain at least one heterologous regulatory sequence responsive to regulation by an exogenous agent. The regulatory sequences are preferably responsive to exogenous agents such as, but not limited to, drugs, hormones, or other gene products. For example, the regulatory sequences, e.g., promoter, preferably are responsive to glucocorticoid receptor-hormone complexes, which, in turn, enhance the level of transcription of a therapeutic peptide or a therapeutic fragment thereof.

[0043] One of ordinary skill in the art will appreciate that each promoter drives transcription, and, therefore, protein expression, differently with respect to time and amount of protein produced. For example, the CMV promoter is characterized as having peak activity shortly after transduction, i.e., about 24 hours after transduction, then quickly tapering off. On the other hand, the RSV promoter's activity increases gradually, reaching peak activity several days after transduction, and maintains a high level of activity for several weeks. Indeed, sustained expression driven by an RSV promoter has been observed in all cell types studied, including, for instance, liver cells, lung cells, spleen cells, diaphragm cells, skeletal muscle cells, and cardiac muscle cells. Thus, a promoter can be selected for use in the invention by matching its particular pattern of activity with the desired pattern and level of expression of a nucleic acid sequence of interest. Alternatively, a hybrid promoter can be constructed which combines the desirable aspects of multiple promoters. For example, a CMV-RSV hybrid promoter combining the CMV promoter's initial rush of activity with the RSV promoter's high maintenance level of activity would be especially preferred for use in many embodiments of the invention. It is also possible to select a promoter with an expression profile that can be manipulated by an investigator.

[0044] A nucleic acid sequence encoding a marker protein, such as green fluorescent protein or luciferase also can be present in the expression construct. Such marker proteins are useful in construction of the expression construct as well as in determining expression construct migration if administered to an organism. Marker proteins also can be used to

determine points of injection in order to efficiently space injections of an expression construct composition to provide a widespread area of treatment, if desired. Alternatively, a nucleic acid sequence encoding a selection factor, which also is useful in vector construction protocols, can be part of the expression construct.

[0045] Negative selection genes can be incorporated into any of the above-described expression constructs. A preferred embodiment is an HSV tk gene cassette (Zjilstra et al., *Nature,* 342: 435 (1989); Mansour et al., *Nature,* 336: 348 (1988); Johnson et al., *Science,* 245: 1234 (1989): Adair et al., *PNAS*, 86: 4574 (1989); Capecchi, *Science,* 244: 1288 (1989)) operably linked to the E2 promoter. The tk expression cassette (or other negative selection expression cassette) is inserted into an adenoviral genome, for example, as a replacement for a substantial deletion of the E3 gene. Other negative selection genes will be apparent to those of skill in the art.

[0046] With respect to promoters, nucleic acid sequences, selectable markers, and the like, located on an expression construct according to the invention, such elements can be present as part of a cassette, either independently or coupled. In the context of the invention, a "cassette" is a particular base sequence that possesses functions, which facilitate subcloning, and recovery of nucleic acid sequences (e.g., one or more restriction sites) or expression (e.g., polyadenylation or splice sites) of particular nucleic acid sequences.

[0047] Construction of a nucleic acid sequence operably linked to regulatory sequences necessary for expression is well within the skill of the art (see, for example, Sambrook et al. (1989), *supra*). With respect to the expression of nucleic acid sequences according to the invention, the ordinary skilled artisan is aware that different genetic signals and processing events control levels of nucleic acids and proteins/peptides in a cell, such as, for instance, transcription, mRNA translation, and post-transcriptional processing. Transcription of DNA into RNA requires a functional promoter, as described herein.

[0048] Protein expression is dependent on the level of RNA transcription that is regulated by DNA signals, and the levels of DNA template. Similarly, translation of mRNA requires, at the very least, an AUG initiation codon, which is usually located within 10 to 100 nucleotides of the 5' end of the message. Sequences flanking the AUG initiator codon have been shown to influence its recognition by eukaryotic ribosomes, with conformity to a perfect Kozak consensus sequence resulting in optimal translation (see, e.g., Kozak, *J. Mol. Biol.*, 196: 947-950 (1987)). Also, successful expression of an exogenous nucleic acid in a cell can require post-translational modification of a resultant protein. Thus, production of a protein can be affected by the efficiency with which DNA (or RNA) is transcribed into mRNA, the efficiency with which mRNA is translated into protein, and the ability of the cell to carry out post-translational modification. These are all factors of which the ordinary

skilled artisan is aware and is capable of manipulating using standard means to achieve the desired end result.

[0049] Along these lines, to optimize protein production, preferably the nucleic acid sequence encoding a protein further comprises a polyadenylation site following the coding region of the nucleic acid sequence. Also, preferably all the proper transcription signals (and translation signals, where appropriate) will be correctly arranged such that the nucleic acid sequence will be properly expressed in the cells into which it is introduced. If desired, the nucleic acid sequence also can incorporate splice sites (i.e., splice acceptor and splice donor sites) to facilitate mRNA production. Moreover, if the nucleic acid sequence encodes a protein or peptide, which is a processed or secreted protein or acts intracellularly, preferably the nucleic acid sequence further comprises the appropriate sequences for processing, secretion, intracellular localization, and the like.

[0050] It will be appreciated that the expression construct can comprise multiple nucleic acid sequences of interest. For example, the expression construct can comprise multiple copies of a nucleic acid sequence encoding a therapeutic factor, each copy operably linked to a different promoter or to identical promoters. Moreover, any nucleic acid sequence encoding a therapeutic factor described herein can be altered from its native form to increase its therapeutic effect. For example, a cytoplasmic form of a therapeutic nucleic acid can be converted to a secreted form by incorporating a signal peptide into the encoded gene product.

[0051] The invention also provides a method of prophylactically or therapeutically treating a mammal for a pathologic state. The method comprises (a) administering to a mammal an expression construct comprising a nucleic acid sequence encoding an AAV IEE and a nucleic acid sequence encoding a therapeutic factor, wherein the expression construct is substantially devoid of AAV ITRs, and (b) administering to the mammal a nucleic acid sequence encoding an AAV Rep protein, such that the nucleic acid sequence encoding the AAV Rep protein is expressed, thereby resulting in the site-specific integration of the nucleic acid sequence encoding a therapeutic factor into a chromosome of a cell of the mammal, the expression of the nucleic acid sequence encoding a therapeutic factor, and the subsequent production of the therapeutic factor to prophylactically or therapeutically treat the mammal for the pathologic state.

[0052] By "prophylactic" is meant the protection, in whole or in part, against a pathologic state. By "therapeutic" is meant the amelioration, in whole or in part, of the pathologic state, itself, and/or the protection, in whole or in part, against further progression of the disease. One of ordinary skill in the art will appreciate that any degree of protection from, or amelioration of, a pathologic state is beneficial to a patient.

[0053] When used for therapeutic purposes, the expression construct of the invention can be purified from a host cell using a variety of conventional purification methods, such as CsCl gradients or chromatography (e.g., ion-exchange chromatography). Such purification techniques are well known and frequently practiced in the art.

[0054] An expression construct of the invention desirably is formulated and administered to a mammal in a composition (i.e., an expression construct composition). Such compositions typically comprise an expression construct and a carrier. Preferably, the carrier is a pharmaceutically (e.g., physiologically) acceptable carrier and can be used within the context of the invention. Such carriers are well known in the art. The choice of carrier will be determined, in part, by the particular site to which the composition is to be administered and the particular method used to administer the adenoviral vector composition.

[0055] Suitable formulations include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain anti-oxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood or intraocular fluid of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The formulations can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials, and can be stored in a freeze-dried (lyophilized) condition requiring only the addition of the sterile liquid carrier, for example, water, immediately prior to use. Extemporaneous solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. When administering an expression construct composition, preferably the pharmaceutically acceptable carrier is a buffered saline solution. More preferably, the expression construct composition for use in the invention is administered in an expression construct composition formulated to protect the expression construct from damage prior to administration. For example, the expression construct composition can be formulated to reduce loss of the expression construct on devices used to prepare, store, or administer the expression construct composition, such as glassware, syringes, or needles. The expression construct composition can be formulated to decrease the light sensitivity and/or temperature sensitivity of the expression construct itself. To this end, the expression construct composition preferably comprises a pharmaceutically acceptable liquid carrier, such as, for example, those described above, and a stabilizing agent selected from the group consisting of polysorbate 80, L-arginine, polyvinylpyrrolidone, trehalose, and combinations thereof (see, e.g., U.S. Patent 6,225,289). Use of such an expression construct composition will extend the shelf life of the expression construct composition, facilitate administration, and increase the effectiveness of the expression construct. In this regard, an expression construct composition also can be formulated to enhance transduction efficiency.

[0056] In addition, the composition of the invention can comprise, or alternatively can be co-administered with, other therapeutic or biologically active agents. By "co-administration" is meant administration before, concurrently with, e.g., in combination with the expression construct in the same formulation or in separate formulations, or after administration of the expression construct as described above. For example, nucleic acid sequences, proteins, and/or other agents useful in the treatment of a particular pathologic state can be present of co-administered with the composition of the invention. Suitable biologically active agents can include, for example, factors that control inflammation, such as ibuprofen or steroids, which can be co-administered to reduce swelling and inflammation associated with administration of the expression construct. Immunosuppressive agents can be co-administered to reduce inappropriate immune responses related to a disorder or the practice of the inventive method. Anti-angiogenic factors, such as soluble growth factor receptors, growth factor antagonists, i.e., angiotensin, and the like also can be co-administered, as well as can be neurotrophic factors. Similarly, vitamins and minerals, anti-oxidants, and micronutrients can be co-administered. Antibiotics, i.e., microbicides and fungicides, can be co-administered to reduce the risk of infection associated with a particular pathologic state. When treating cancer, other anticancer compounds can be used in conjunction with the composition of the invention and can include, but are not limited to, all of the known anticancer compounds approved for marketing in the United States and those that will become approved in the future. See, for example, Table 1 and Table 2 of Boyd, Current Therapy in Oncology, Section 1. Introduction to Cancer Therapy (J.E. Niederhuber, ed.), Chapter 2, by B.C. Decker, Inc., Philadelphia, 1993, pp. 11-22. More particularly, such other anticancer compounds can include doxorubicin, bleomycin, vincristine, vinblastine, VP-16, VW-26, cisplatin, carboplatin, procarbazine, and taxol for solid tumors in general; alkylating agents, such as BCNU, CCNU, methyl-CCNU and DTIC, for brain or kidney cancers; and antimetabolites, such as 5-FU and methotrexate, for colon cancer.

[0057] The pathologic state can be any pathologic state. For example, the pathologic state can be a disorder caused by an increased or decreased level of a particular gene product(s). By "increased level" is meant a level above that which is considered normal. Similarly, by "decreased level" is meant a level below that which is considered normal. Many cancers result from an increased level of an oncogene, or, alternatively, a decreased level of a tumor suppressor gene. Accordingly, the pathologic state preferably is cancer, and the nucleic acid sequence preferably encodes a therapeutic factor, which is toxic to one or more different cancer cell types.

[0058] The pathologic state can be any type of cancer. Cancers can include lung cancer, colon cancer, renal cancer, anal cancer, bile duct cancer, bladder cancer, bone cancer, brain

cancer, spinal chord cancer, breast cancer, cervical cancer, lymphoma, endometrial cancer, esophageal cancer, gallbladder cancer, gastrointestinal cancer, laryngeal cancer, leukemia, liver cancer, multiple myeloma, neuroblastoma, ovarian cancer, pancreatic cancer, prostatic cancer, retinoblastoma, skin cancer (e.g., melanoma and non-melanoma), stomach cancer, testicular cancer, thymus cancer, and thyroid cancer, as well as other carcinomas and sarcomas.

[0059] Other pathologic states are also contemplated in the context of the invention. For example, the pathologic state can be an inflammatory disease (e.g., arthritis), a neurodegenerative disease, a disease of an organ which is attributed to the presence of the increased or decreased level of a particular gene product(s), or any other pathologic state for which the site-specific integration and subsequent expression of a nucleic acid sequence encoding a therapeutic factor will treat or prevent a particular pathologic state.

[0060] Suitable methods, both invasive and noninvasive methods, of directly administering an expression construct (e.g., an expression construct composition) are available. Although more than one route can be used for administration, a particular route can provide a more immediate and more effective reaction than another route. The inventive method is not dependent on the mode of administering the expression construct composition to a mammal, preferably a human, to achieve the desired effect. As such, any route of administration is appropriate so long as the expression construct contacts and enters a cell within which integration is achievable. The composition can be appropriately formulated and administered in the form of a local injection, lotion, ointment, implant, or the like. The expression construct composition can be applied, for example, topically, intratumoraly, or peritumoraly. The expression construct composition can be administered through multiple applications and/or multiple routes to ensure sufficient exposure of cells to the expression construct composition.

[0061] The expression construct is preferably formulated into a composition prior to administration and is administered as soon as possible after it has been determined that an animal, such as a mammal, specifically a human, is at risk for a particular pathologic state (prophylactic treatment) or has begun to develop the pathologic state (therapeutic treatment). Treatment will depend, in part, upon the particular therapeutic factor expressed from the nucleic acid sequence, the route of administration, and the cause and extent, if any, of the pathologic state.

[0062] The expression construct can be administered using invasive procedures, such as, for instance, local injection (e.g., intratumoral injection). Intratumoral injections involve the administration of the expression construct composition directly into a tumor cell(s), which desirably selectively allow for expression construct replication. Pharmaceutically acceptable carriers for injectable compositions are well known to those of ordinary skill in

the art (see Pharmaceutics and Pharmacy Practice, J.B. Lippincott Co., Philadelphia, PA, Banker and Chalmers, eds., pages 238-250 (1982), and ASHP Handbook on Injectable Drugs, Toissel, 4th ed., pages 622-630 (1986)).

[0063] The expression construct can be non-invasively administered to a mammal. For instance, if multiple surgeries have been performed, the mammal displays low tolerance to anesthetic, or other disorders exist, topical administration of the expression construct composition may be most appropriate. Topical formulations are well known to those of skill in the art. An expression construct composition also can be administered non-invasively using a needleless injection device, such as the Biojector 2000 Needle-Free Injection Management System® available from Bioject, Inc.

[0064] The expression construct is preferably present in or on a device that allows controlled or sustained release, such as a biocompatible polymeric matrix, meshwork, mechanical reservoir, or mechanical implant. Implants (see, e.g., U.S. Patents 5,443,505, 4,853,224 and 4,997,652) and devices (see, e.g., U.S. Patents 5,554,187, 4,863,457, 5,098,443 and 5,725,493), such as an implantable device, e.g., a mechanical reservoir or an implant or a device comprised of a polymeric composition, are particularly useful for the administration of the expression construct composition. The expression construct also can be administered in the form of a sustained-release formulation (see, e.g., U.S. Patent 5,378,475) comprising, for example, gelatin, chondroitin sulfate, a polyphosphoester, such as bis-2-hydroxyethyl-terephthalate (BHET), or a polylactic-glycolic acid.

[0065] When administering the expression construct, the appropriate dosage and route of administration can be selected to minimize loss of the expression construct or inactivation of the expression construct due to a host's immune system. For example, for contacting cells *in vivo*, it can be advantageous to administer, to a mammal being treated, an immunosuppressive agent (e.g., cyclophosphamide or FK506) or monoclonal antibody that can block a T cell receptor, prior to performing the inventive method. Prior administration of an immunosuppressive agent or monoclonal antibody can serve to decrease the amount of expression construct cleared by the immune system of the mammal.

[0066] When practiced *in vivo*, any suitable organs or tissues or component cells can be targeted for expression construct delivery. Preferably, the organs/tissues/cells employed are of the circulatory system (i.e., heart, blood vessels or blood), respiratory system (i.e., nose, pharynx, larynx, trachea, bronchi, bronchioles, lungs), gastrointestinal system (i.e., mouth, pharynx, esophagus, stomach, intestines, salivary glands, pancreas, liver, gallbladder), urinary system (i.e., kidneys, ureters, urinary bladder, urethra), nervous system (i.e., brain and spinal cord, and special sense organs such as the eye) and integumentary system (i.e., skin). Even more preferably, the cells being targeted are selected from the group consisting of heart, blood vessel, lung, liver, gallbladder, urinary bladder, and eye cells.

[0067] The dose of expression construct administered to a mammal, particularly a human, in accordance with the invention should be in an amount sufficient to treat prophylactically or therapeutically a mammal for a pathologic state. Dosage will depend upon a variety of factors, including the age, species, the pathology in question, and condition or disease state. Dosage also depends on the nucleic acid sequences contained in the expression construct, as well as the amount of tissue about to be affected or actually affected by the disease. The size of the dose also will be determined by the route, timing, and frequency of administration as well as the existence, nature, and extent of any adverse side effects that might accompany the administration of a particular expression construct and the desired physiological effect. It will be appreciated by one of ordinary skill in the art that various conditions or disease states, in particular, chronic conditions or disease states, may require prolonged treatment involving multiple administrations.

[0068] Suitable doses and dosage regimens can be determined by conventional range-finding techniques known to those of ordinary skill in the art. When administering an expression construct, preferably about 10^6 particles to about 10^{12} particles (e.g., naked DNA particles, plasmid particles, viral particles) are delivered to the diseased tissue. In other words, a composition of the expression construct can be administered that comprises an expression construct concentration of about 10^6 particles/ml to about 10^{12} particles/ml (including all integers within the range of about 10^6 particles/ml to about 10^{12} particles/ml), preferably about 10^{10} particles/ml to about 10^{12} particles/ml. Typically, about 0.1 μl to about 100 μl of such an expression construct composition to each affected tissue. Of course, other routes of administration may require smaller or larger doses to achieve a therapeutic effect. Any necessary variations in dosages and routes of administration can be determined by the ordinarily skilled artisan using routine techniques known in the art.

[0069] In some embodiments, it is advantageous to administer two or more (i.e., multiple) doses of the expression construct. The invention provides for multiple applications of the expression construct in order to achieve sufficient integration, thereby prophylactically or therapeutically treating a particular disease state. For example, at least two applications of an expression construct can be administered to the same tissue. Preferably, the cell(s) is contacted with two applications or more of the expression construct via direct administration to the desired tissue within about 30 days or more. More preferably, two or more applications are administered to cells of the same tissue within about 90 days or more. However, three, four, five, six, or more doses can be administered in any time frame (e.g., 2, 7, 10, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 85 or more days between doses) so long as the desired therapeutic effect is achieved. However, because the expression constructs of the invention are extremely efficient in integration, multiple applications are likely unnecessary.

[0070] The expression construct can be introduced *ex vivo* into cells, previously removed from the mammal, especially a human, and exposed to the expression construct, although this is less preferred. Such transduced autologous or homologous host cells, reintroduced into the mammal (e.g., human), will express directly the nucleic acid sequences contained therein *in vivo* following initiation of DNA replication. One *ex vivo* therapeutic option involves the encapsidation of infected cells into a biocompatible capsule, which can be implanted into a particular tissue. Such cells need not be isolated from the patient, but can instead be isolated from another individual and implanted into the patient.

[0071] The following examples further illustrate the invention but, of course, should not be construed as in any way limiting its scope.

EXAMPLE 1

[0072] This example illustrates that a *cis*-acting element, other than AAV ITRs, functions as an integration efficiency element.

[0073] Two recombinant AAV (rAAV) plasmids were constructed from adeno-associated virus 2 (pRepGFP and pGFPCap). pRepGFP was constructed to contain, from 5' to 3', an AAV ITR, the Rep ORF operably linked to a nucleic acid sequence encoding an AAV IEE, a GFP transgene operably linked to a CMV promoter, and a second AAV ITR. pGFPCap was constructed to contain, from 5' to 3', an AAV ITR, a nucleic acid sequence encoding an AAV IEE, a GFP transgene operably linked to a CMV promoter, the second half of the wt AAV genome including the AAV Cap ORF, and a second AAV ITR. In addition, pTRUF2, which is a well-known rAAV plasmid, was utilized to analyze the integration efficiency of an expression construct which lacks a nucleic acid sequence encoding an AAV IEE. In that respect, pTRUF2 contains, from 5' to 3', an AAV ITR, a GFP transgene and a neomycin resistance gene operably linked to distinct promoters, and a second AAV ITR (see, e.g., Zolotukhin et al., *J. Virol.*, 70:4646-4654 (1996)).

[0074] HeLa cells were co-transfected with pSub201 and either of pRepGFP, pGFPCap, or pTRUF2. pSub201 is a plasmid carrying the entire wtAAV genome. It is co-transfected with a corresponding rAAV plasmid to provide Rep *in trans*. Transfected cells were sorted by fluorescence-activated cell sorting (FACS) after 48 hours (Beckman-Coulter Altra cell sorter). The sorted cells were plated on a 96 well plate with 1 cell/well, and were allowed to grow for approximately 6 weeks. After this time, whole cell DNA was isolated from HeLa cells using a standard protocol (see, e.g., Miller et al., *Nucleic Acids Res.*, 16:1215 (1988)). PCR and Southern Blots were performed to screen for GFP expression to determine if integration had occurred. The results of these experiments are summarized in Table 1.

Table 1

Plasmid	Integration Efficiency Rate
pRepGFP	27%
pGFPCap	5%
pSub201	12%
pTRUF2	< 1%

[0075] As indicated by the results set forth in Table 1, pRepGFP integrated at a rate of approximately 27%, substantially higher than the 12% integration efficiency observed with pSub201. Also, the integration efficiency of pGFPCap was 5%, which is lower than the 12% integration efficiency observed with pSub201 but higher than that observed with pTRUF2. These results demonstrate that the nucleic acid sequence encoding an AAV IEE (which was present in pRepGFP and pGFPCap, but absent in pTRUF2) enables efficient transgene integration.

EXAMPLE 2

[0076] This example demonstrates that the deletion of AAV ITRs in a recombinant AAV plasmid does not influence the efficiency of Rep mediated site-specific integration.

[0077] Plasmids were constructed as described in Example 1 but with complete deletions of AAV ITRs. HeLa cells were co-transfected with pSub201 and either of pRepGFP(ITR-), pGFPCap(ITR-), or pTRUF2(ITR-). Transfected cells were sorted, plated, and grown as described in Example 1. Whole cell DNA was isolated from HeLa cells as described in Example 1. PCR and Southern Blots were performed to screen for GFP expression to determine if site-specific integration had occurred. The results are summarized in Table 2.

Table 2

Plasmid	Integration Efficiency Rate
pRepGFP(ITR-)	27%
pGFPCap(ITR-)	5%
pSub201(ITR-)	12%
pTRUF2(ITR-)	< 1%

[0078] As indicated by the results set forth in Table 2, expression constructs lacking AAV ITRs integrated into the host cell genome by Rep-mediated site-specific integration at a similar efficiency to their ITR-containing counterparts. These results demonstrate that

AAV ITRs are not necessary for enabling site-specific integration, and, moreover, that a nucleic acid sequence encoding an AAV IEE is sufficient for directing site-specific integration of a transgene (e.g., in the absence of AAV ITRs).

EXAMPLE 3

[0079] This example further demonstrates that AAV-ITR elements are not required for AAV site-specific integration.

[0080] HeLa cells were transfected with a pAAV/Ad plasmid (pRepCap(itr-)) (Samulski et al., *J. Virol., 63,* 3822-3828 (1989)) and a pCMV-GFB plasmid (Philpott et al., *Proc. Natl. Acad. Sci., 99(19),* 12381-12385 (2002)). pRepCap(itr-) contains wild-type AAV sequences with both flanking ITR elements deleted. pCMV-GFB expresses GFP and does not contain an AAV sequence. As a control, HeLa cells were transfected with a pSub201 plasmid (pRepCap(itr+)) (containing wild-type AAV sequences including both flanking ITR elements) and a pCMV-GFB plasmid.

[0081] Transfected cells were sorted by fluorescence-activated cell sorting (FACS) after 48 hours (Beckman-Coulter Altra cell sorter). The sorted cells were plated on a 96 well plate with 1 cell/well, and the cells were allowed to grow for approximately 6 weeks. After this time, genomic DNA was isolated from HeLa cells using a standard protocol (see, e.g., Miller et al., *supra*). PCR and Southern Blots were performed to screen for GFP expression to determine if integration had occurred and for the disruption of the AAVS1 genomic locus. The results are summarized in Table 3.

Table 3

Plasmids	Number of tested clones	Percentage of clones with Rep	Percentage of clones with AAVS1 disruption
pRepCap(itr+) + pCMV-GFB	48	12	33
pRepCap(itr-) + pCMV-GFB	49	22	45

[0082] As indicated above, the wild-type ITR-containing plasmid, pRepCap(itr+), had a Rep integration efficiency of 12%. AAVSI disruptions were found in a higher proportion of cell lines (33%) than those which retained the Rep DNA sequence.

[0083] In cells containing pRepCap(itr-) plasmid, Rep was found in 22% of the cell lines tested, and in most cases Rep was integrated into AAVS1. Furthermore, a large percentage (45%) of pRepCap(itr-) transfected cell lines had AAVSI genomic fragment disruptions.

[0084] From these results, it is apparent that plasmid constructs that lack the AAV ITR elements can serve as substrate DNA for rep-mediated site-specific integration into human chromosome 19 at the AAVSI site. Furthermore, it appears that higher levels of AAVS1 disruption than substrate DNA integrations are found in cell lines established from either ITR+ or ITR- substrate plasmids.

[0085] This results of this experiment indicate that ITR elements are not required for site-specific integration into AAVS1, and that pRepCap(itr-) is an independent substrate for Rep-mediated site-specific integration into AAVS1.

EXAMPLE 4

[0086] This example also demonstrates that AAV-ITR elements are not required for AAV site-specific integration.

[0087] It is commonly observed that when AAV integrates into HeLa AAVS1 sites, the resulting disruptions are variable in AAVS1 restriction fragment length and fragment band intensity. This phenomena may be explained by the influence of several factors on the character of AAVS1 integrants. For example, these factors may include the instability of the AAVS1 integration site (particularly in the presence of Rep), the aneuploidy of HeLa cells, and the imprecise alteration of the sequence of the AAVS1 site by the deletion-insertion mechanism of Rep-mediated AAVS1 integrations.

[0088] HeLa cells were cotransfected with a Rep-expressing plasmid to mediate the integration event and a GFP plasmid for FACS sorting transduced cells. Specifically, HeLa cells were transfected with the Rep-expressing plasmid, pRepCap(itr-) and pGFPCap (both described above). pGFPCap contains the first 7% of the AAV genome, followed by a GFP transgene and AAV Cap sequence.

[0089] Transfected cells were sorted by fluorescence-activated cell sorting (FACS) after 48 hours (Beckman-Coulter Altra cell sorter). The sorted cells were plated on a 96 well plate with 1 cell/well, and the cells were allowed to grow for approximately 6 weeks. After this time, genomic DNA was isolated from HeLa cells using a standard protocol (see, e.g., Miller et al., *supra*). PCR and Southern Blots were performed to screen for GFP expression to determine if integration had occurred and for the disruption of the AAVS1 genomic locus.

[0090] Approximately 6% of the 78 tested cell lines contained integrated pGFPCap DNA. When the cell lines were screened for genomic alterations of the AAVS1 site of chromosome 19, 54% of the clones were found to have AAVS1 site disruptions. As observed in Example 3, it appears that AAVS1 disruptions occurred at much higher frequencies than pGFPCap integration events. This observation may be explained by one or more of the following: Rep-mediated disruption of AAVS1 sites can occur in the

absence of integration; integration events are occurring at AAVS1 sites but the integrated DNA is unstable, or DNA elements other than pGFPCap are integrating at the AAVS1 site.

[0091] 14% of the cell lines contained Rep DNA that in most cases was site-specifically integrated. If the ITRs were required for AAV site-specific integration, such a large percentage of integrants would not have been observed. Since it was possible that a recombination event between the pGFPCap and pRepCap(itr-) plasmids to form wild-type AAV comprising ITR elements (the plasmids share significant homology at both the 5' and 3' ends of their respective AAV regions) was responsible for the large number of Rep integrants, PCR analysis of genomic DNA was performed using one primer complementary to the pRepCap(itr-) Rep sequence and one primer complementary to the pGFPCap ITR sequence. A PCR product of about 700 bp would indicate a wild-type recombination event, whereas neither of the two plasmids separately would serve as a substrate for the PCR primer pair. A PCR product was not obtained from any of the pGFPCap-pRepCap(itr-) co-transfected cell lines, which suggested that recombination between pRepCap(itr-) and pGFPCap to form a wild-type AAV plasmid was unlikely to have accounted for the Rep integrants. This result supports the finding that pRepCap(itr-) is an independent substrate for Rep-mediated site-specific integration into AAVS1 (i.e., ITR elements are not required for Rep-mediated site specific integration into AAVS1).

EXAMPLE 5

[0092] This example demonstrates that ITR elements influence the boundaries of integration substrates.

[0093] As described in Example 3, genomic DNA was isolated from the pRepCap(itr-) and the pRepCap(itr+) transfected clones (each of which was co-transfected with pCMV-GFB). The DNA was digested with *Eco R1* and separated on 1% agarose gels. The resulting DNA was transferred to nylon membranes and hybridized to a ^{32}P-labeled probe of the pRepCap(itr-) plasmid backbone sequence. The results of this analysis are summarized in Table 4.

Table 4

Plasmids	Number of tested clones	Percentage of clones with plasmid backbone	Percentage site-specificity
pRepCap(itr+) + pCMV-GFB	45	23	90
pRepCap(itr-) + pCMV-GFB	49	24	91

[0094] The Southern blot of genomic DNA from pRepCap(itr-)-transfected cells probed with plasmid backbone showed 24% of the cell lines contained plasmid backbone. Additionally, 91% of the Rep-positive clones were positive for the plasmid backbone. Based on these results, it appears that the entire plasmid sequence was the substrate for AAVS1 integration in cell lines derived from a pRepCap(itr-) transduction of HeLa cells.

[0095] The Southern blot of genomic DNA from pRepCap(itr+)-transfected cells probed with plasmid backbone resulted in a more complex pattern of AAVS1 integration. 27% of cell lines tested had at least part of the pRepCap(itr+) sequence integrated. Rep was site-specifically integrated into 12% of the cell lines (as discussed in Example 3). 23% of the cell lines were shown to contain plasmid integrants.

[0096] The three types of integrants observed following co-transfection of pRepCap(itr+) and pCMV-GFB may be explained as follows. The terminal hairpin structure of the AAV ITR serves as the viral origin of replication. In a double-stranded plasmid substrate, such as in pRepCap(itr+), duplex cruciform structures can be generated at each ITR, with each cruciform being a substrate for Rep binding and nicking. If nicking at an ITR cruciform occurs, it may result in defining the segment of pRepCap(itr+) element that is able to undergo site-specific integration. If this occurs, three distinct integration substrates may result: RepCap flanked by ITR elements, plasmid backbone flanked by ITR elements, or both RepCap and plasmid backbone flanked by ITR elements. In contrast, pRepCap(itr-) lacks the complexity and function of the ITR elements, and all integrants of pRepCap(itr-) should include the entire pRepCap(itr-) DNA sequence.

[0097] The overall integration efficiency of ITR-containing or ITR-deleted constructs (27% and 24%, respectively) was similar. Therefore, the AAV ITRs made a minimal contribution toward integration efficiency. However, the presence of the AAV ITR elements can act to generate integration boundaries when using a plasmid integration substrate. Presumably, Rep binding to the Rep-binding elements (RBEs) and nicking at the terminal resolution site (trs) of the ITR generates potential 5' and 3' boundaries of the integration substrate. In a wild-type viral infection, boundary definition is naturally present due to the ITR hairpin structure present at the ends of the linear, single-stranded viral genome.

EXAMPLE 6

[0098] This example demonstrates that the only *cis*-element that is necessary and sufficient for site-specific integration at AAVS1 is present in the sequence comprising the AAV integration efficiency element (IEE) and p5 promoter region. The p5 promoter region of AAV overlaps with the IEE.

[0099] The plasmid pAd-p5CAT containing the p5 promoter region and AAVIEE upstream of a chloramphenicol acetyl transferase (CAT) reporter gene was constructed (Philpott et al., *Proc. Natl. Acad. Sci., supra*). Promoter activity was confirmed with this construct in transient transfection assays. In HeLa cells, the p5 promoter was able to mediate expression of CAT levels comparable to the CMV promoter and was vulnerable to *trans* repression by Rep.

[00100] To determine the function of the construct in an integration assay, HeLa cells were co-transfected with pAd-p5CAT and the Rep-expressing plasmid pT7-Rep (Philpott et al., *J. Virol., 76,* 5411-5421 (2002)). The pT7-Rep plasmid expressed GFP, which allowed for FACS sorting 24 hours post-transfection. Cell lines were established as previously described, and each cell line was harvested for genomic DNA. Genomic DNA was digested with *Eco R1*, and a Southern blot analysis was performed probing for the presence of CAT or for the presence of AAVS1 disruptions. The results are summarized in Table 5.

Table 5

Plasmids	Number of tested clones	Percentage of clones with plasmid backbone	Percentage with AAVS1 disruption	Percentage site-specificity
pT7-Rep + pAd-p5CAT	45	22	29	100
pAd-p5CAT	49	0	0	0

[00101] Out of 45 cell lines tested, 29% contained AAVS1 disruptions, and CAT gene integrations were found in 22% of cell lines. In all cases, pAd-p5CAT DNA co-migrated with the AAVS1 probe (i.e. 100% site specificity to the AAVS1 site).

[00102] As a negative control, HeLa cells were transfected with pAd-p5CAT alone. In the absence of a Rep-expressing plasmid, the CAT transgene was unable to integrate, and this result confirms that the targeted integration of a substrate is Rep-dependent.

[00103] The results of this experiment indicate that the sequence comprising the p5 promoter region and IEE is the only AAV element required in *cis* to mediate site-specific integration of a substrate DNA through Rep-dependent integration into AAVS1.

[00104] All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

[00105] The use of the terms "a" and "an" and "the" and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (i.e., meaning "including, but not limited to,") unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[00106] Preferred embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

WHAT IS CLAIMED IS:

1. An expression construct comprising a nucleic acid sequence encoding an adeno-associated virus integration efficiency element (AAV IEE), wherein the expression construct is substantially devoid of AAV inverted terminal repeats (AAV ITRs), and wherein the expression construct site-specifically integrates into a host cell chromosome when provided to the host cell in conjunction with an AAV Rep protein.

2. The expression construct of claim 1, wherein the Rep protein comprises a long form of Rep.

3. The expression construct of claim 1 or 2, wherein the expression construct further comprises a nucleic acid sequence of interest.

4. The expression construct of claim 3, wherein the nucleic acid sequence of interest encodes a protein.

5. The expression construct of claim 4, wherein the protein is a therapeutic factor.

6. The expression construct of claim 5, wherein the nucleic acid sequence of interest encoding a therapeutic factor is positioned downstream of the AAV IEE.

7. The expression construct of claim 5 or 6, wherein the therapeutic factor is useful to prophylactically or therapeutically treat a mammal for a pathologic state.

8. The expression construct of any of claims 1-7, wherein the expression construct is a viral vector.

9. The expression construct of claim 8, wherein the viral vector is an adenoviral vector.

10. The expression construct of claim 9, wherein the adenoviral vector is deficient in one or more replication-essential gene functions.

11. A composition comprising the expression construct of any of claims 1-10 and a carrier.

12. A host cell comprising the expression construct of any of claims 1-10.

13. A host cell comprising the expression construct of any of claims 1-10.

14. The host cell of claim 13, wherein the host cell is propagated to produce a stable cell line.

15. A method of integrating a nucleic acid sequence of interest into a host cell chromosome comprising:

(a) providing to a host cell an expression construct comprising a nucleic acid sequence encoding an adeno-associated virus integration efficiency element (AAV IEE) and a nucleic acid sequence of interest, wherein the expression construct is substantially devoid of AAV inverted terminal repeats (AAV ITRs), and

(b) providing to the host cell a nucleic acid sequence encoding an AAV Rep protein,

such that the nucleic acid sequence encoding the AAV Rep protein is expressed, thereby resulting in the site-specific integration of the nucleic acid sequence of interest into a chromosome contained in the host cell.

16. The method of claim 15, wherein the nucleic acid sequence of interest encodes a protein.

17. The method of claim 16, wherein the protein is a therapeutic factor.

18. The method of claim 17, wherein the therapeutic factor is useful to prophylactically or therapeutically treat a mammal for a pathologic state.

19. The expression construct of any of claims 15-18, wherein the AAV Rep protein is provided to the host cell *in trans*.

20. The expression construct of any of claims 15-18, wherein the AAV Rep protein is provided to the host cell in the expression construct of (a).

21. The method of any of claims 15-20, wherein the expression construct is a viral vector.

22. The method of claim 21, wherein the viral vector is an adenoviral vector.

23. The method of claim 22, wherein the adenoviral vector is deficient in one or more replication-essential gene functions.

24. The method of claim 23, wherein the host cell is propagated to create a stable cell line.

25. A method of prophylactically or therapeutically treating a mammal for a pathologic state comprising:

(a) administering to a mammal an expression construct comprising a nucleic acid sequence encoding an adeno-associated virus integration efficiency element (AAV IEE) and a nucleic acid sequence encoding a therapeutic factor, wherein the expression construct is substantially devoid of AAV inverted terminal repeats (AAV ITRs), and

(b) administering to the mammal a nucleic acid sequence encoding an AAV Rep protein,

such that the nucleic acid sequence encoding the AAV Rep protein is expressed, thereby resulting in the site-specific integration of the nucleic acid sequence encoding a therapeutic factor into a chromosome of a host cell of the mammal, expression of the nucleic acid sequence encoding a therapeutic factor, and subsequent production of the therapeutic factor to prophylactically or therapeutically treat the mammal for the pathologic state.

26. The expression construct of claim 25, wherein the AAV Rep protein is provided to the host cell *in trans*.

27. The expression construct of claim 25, wherein the AAV Rep protein is provided to the host cell in the expression construct of (a).

28. The method of any of claims 25-27, wherein the expression construct is a viral vector.

29. The method of claim 28, wherein the viral vector is an adenoviral vector.

30. The method of claim 29, wherein the adenoviral vector is deficient in one or more replication-essential gene functions.

31. The method of any of claims 25-30, wherein the pathologic state is cancer.

CGACATTTTGCGACACCATGTGGTCACGCTGGGTATTTAAGCCCGAGTGAGCACGCAGGG 60
TCTCCATTTTGAAGCGGGAGGTTTGAACGCGCAGCCGCCATG 102
(SEQ ID NO:1)

Fig. 1

```
AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACC 50
AAV_1_complete           TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCCTGCGGACC 50
AAV_3_complete           TTGGCCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGGCCTGGCGACC 50
AAV_4_complete           TTGGCCACTCCCTCTATGCGCGCTCGCTCACTCACTCGGCCCTGGAGACC 50
AAV_5_partial            ----------------------------------------------------

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           AAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGC 100
AAV_1_complete           AAAGGTCCGCAGACGGCAGAGCTCTGCTCTGCCGGCCCCACCGAGCGAGC 100
AAV_3_complete           AAAGGTCGCCAGACGGACGTGCTTTGCACGTCCGGCCCCACCGAGCGAGC 100
AAV_4_complete           AAAGGTCTCCAGACTGCCGGCCTCTGGCCGGCAGGGCCGAGTGAGTGAGC 100
AAV_5_partial            ----------------------------------------------------

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           GAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGG---------- 139
AAV_1_complete           GAGCGCGCAGAGAGGGAGTGGGCAACTCCATCACTAGGG---------- 139
AAV_3_complete           GAGTGCGCATAGAGGGAGTGGCCAACTCCATCACTAGAG---------- 139
AAV_4_complete           GAGCGCGCATAGAGGGAGTGGCCAACTCCATCATCTAGGTTTGCCCACTG 150
AAV_5_partial            ---CGCGACAGGGGGGAGAGTGCCACACTCTCAAGCAAG---------- 36

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           --------GTTCCTGGAGGGGTGGAGTCGTGACGTGAATTACGTCATAGG 181
AAV_1_complete           --------GTAATCGCGAAGCGCCTCCCACGCTGCCGCGTCAGCGCTGAC 181
AAV_3_complete           --------GTATGGCAGTGACGTAACGCGAAGCGCGCGAAGCGAGACCAC 181
AAV_4_complete           ACGTCAATGTGACGTCCTAGGGTTAGGGAGGTCCCTGTATTAGCAGTCAC 200
AAV_5_partial            -----GGGGTTTTGTAAGCAGTGATGTCATAATGATGTAATGCTTATTGT 81

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           GTTAGGGAGGTCCTGTATTAGAGGTCACGTGAGTG---------------- 216
AAV_1_complete           GTAAATTACGTCATAGGGGAGTGGTCCTGTATTAGCTGTCACGTGA----- 227
AAV_3_complete           GCCTACCAG---CTGCGTCAGCAGTCAGGTGACCC---------------- 213
AAV_4_complete           GTGAGTGTCGTATTTCGCGGAGCGTAGCGGAGCGCATACCAAGCTGCCAC 250
AAV_5_partial            CACGCGATAGTTAATGATTAACAGTCATGTGATGTGTTTTATCCAA----- 127

AAV_IEE_from_serotype_2  ----------------------CGACATTTTGCGACACCATGTGGTCAC 27
AAV_6_complete           ------------------TTTTGCGACATTTTGCGACACCATGTGGTCAC 248
AAV_1_complete           --------------GTGCTTTTGCGACATTTTGCGACACCACGTGGCCAT 263
AAV_3_complete           ------------------TTTTGCGACAGTTTGCGACACCACGTGGCCGC 245
AAV_4_complete           GTCACAGCCACGTGGTCCGTTTGCGACAGTTTGCGACACCATGTGGTCA- 299
AAV_5_partial            ------------TAGGAAGAAAGCGCGCGTATGAGTTCTCGCGAGACTTC 165
                                             **     * ** *     *  * *

AAV_IEE_from_serotype_2  GCTGGGTATTTAAGCCCGAGTGAGCACGC-AGGGTCTCCATTTTGAAGCG 76
AAV_6_complete           GCTGGGTATTTAAGCCCGAGTGAGCACGC-AGGGTCTCCATTTTGAAGCG 297
AAV_1_complete           TTAGGGTATATATGGCCGAGTGAGCGAGC-AGGATCTCCATTTTGAC-CG 311
AAV_3_complete           TGAGGGTATATATTCTCGAGTGAGCGAACCAGGAGCTCCATTTTGAC-CG 294
AAV_4_complete           GGAGGGTATATAACCGCGAGTGAGCCAGCGAGGAGCTCCATTTTGCC-CG 348
AAV_5_partial            -CGGGGTATAAAAGACCGAGTGAACGAGCCCGCCGCCATTCTTTGCTCTG 214
                            ******  *     ******* *    *  *    *      ****     *

AAV_IEE_from_serotype_2  GGAGGTTTGAACGCGCA-GCCGCCATG---------------------- 102
AAV_6_complete           GGAGGTTTGAACGCGCA-GC-GCCATGCCGGGGGTTTTACGAGATTGTGAT 345
AAV_1_complete           CGAAATTTGAACGAGCA-GCAGCCATGCCGGGCTTCTACGAGATCGTGAT 360
```

Fig. 2

```
AAV_3_complete                  CGAAATTTGAACGAGCA-GCAGCCATGCCGGGGTTCTACGAGATTGTCCT 343
AAV_4_complete                  CGAATTTTGAACGAGCA-GCAGCCATGCCGGGGTTCTACGAGATCGTGCT 397
AAV_5_partial                   GACTGCTAGAGGACCCTCGCTGCCATGGCTACCTTCTATGAAGTCATTGT 264
                                      * **     *  ** ******

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  TAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCGGCATTTCTGACAGCT 395
AAV_1_complete                  CAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTGACTCGT 410
AAV_3_complete                  GAAGGTCCCGAGTGACCTGGACGAGCGCCTGCCGGGCATTTCTAACTCGT 393
AAV_4_complete                  GAAGGTGCCCAGCGACCTGGACGAGCACCTGCCCGGCATTTCTGACTCTT 447
AAV_5_partial                   TCGCGTCCCATTTGACGTGGAGGAACATCTGCCTGGAATTTCTGACAGCT 314

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  TTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTCTGAC 445
AAV_1_complete                  TTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCCGGATTCTGAC 460
AAV_3_complete                  TTGTTAACTGGGTGGCCGAGAAGGAATGGGACGTGCCGCCGGATTCTGAC 443
AAV_4_complete                  TTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCGCCGGATTCTGAC 497
AAV_5_partial                   TTGTGGACTGGGTAACTGGTCAAATTTGGGAGCTGCCTCCAGAGTCAGAT 364

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  ATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCT 495
AAV_1_complete                  ATGGATCTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCT 510
AAV_3_complete                  ATGGATCCGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAAAAGCT 493
AAV_4_complete                  ATGGACTTGAATCTGATTGAGCAGGCACCCCTGACCGTGGCCGAAAAGCT 547
AAV_5_partial                   TTAAATTTGACTCTGGTTGAACAGCCTCAGTTGACGGTGGCTGATAGAAT 414

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  GCAGCGCGACTTCCTGGTCCAGTGGCGCCGCGTGAGTAAGGCCCCGGAGG 545
AAV_1_complete                  GCAGCGCGACTTCCTGGTCCAATGGCGCCGCGTGAGTAAGGCCCCGGAGG 560
AAV_3_complete                  TCAGCGCGAGTTCCTGGTGGAGTGGCGCCGCGTGAGTAAGGCCCCGGAGG 543
AAV_4_complete                  GCAACGCGAGTTCCTGGTCGAGTGGCGCCGCGTGAGTAAGGCCCCGGAGG 597
AAV_5_partial                   TCGCCGCGTGTTCCTGTACGAGTGGAACAAATTTTCCAAG---CAGGAGT 461

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  CCCTCTTCTTTGTTCAGTTCGAGAAGGGCGAGTCCTACTTCCACCTCCAT 595
AAV_1_complete                  CCCTCTTCTTTGTTCAGTTCGAGAAGGGCGAGTCCTACTTCCACCTCCAT 610
AAV_3_complete                  CCCTCTTTTTTGTCCAGTTCGAAAAGGGGGAGACCTACTTCCACCTGCAC 593
AAV_4_complete                  CCCTCTTCTTTGTCCAGTTCGAGAAGGGGGACAGCTACTTCCACCTGCAC 647
AAV_5_partial                   CCAAATTCTTTGTGCAGTTTGAAAAGGGATCTGAATATTTTCATCTGCAC 511

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  ATTCTGGTGGAGACCACGGGGGTCAAATCCATGGTGCTGGGCCGCTTCCT 645
AAV_1_complete                  ATTCTGGTGGAGACCACGGGGGTCAAATCCATGGTGCTGGGCCGCTTCCT 660
AAV_3_complete                  GTGCTGATTGAGACCATCGGGGTCAAATCCATGGTGGTCGGCCGCTACGT 643
AAV_4_complete                  ATCCTGGTGGAGACCGTGGGCGTCAAATCCATGGTGGTGGGCCGCTACGT 697
AAV_5_partial                   ACGCTTGTGGAGACCTCCGGCATCTCTTCCATGGTCCTCGGCCGCTACGT 561

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  GAGTCAGATTAGGGACAAGCTGGTGCAGACCATCTACCGCGGGATCGAGC 695
AAV_1_complete                  GAGTCAGATTAGGGACAAGCTGGTGCAGACCATCTACCGCGGGATCGAGC 710
AAV_3_complete                  GAGCCAGATTAAAGAGAAGCTGGTGACCCGCATCTACCGCGGGGTCGAGC 693
AAV_4_complete                  GAGCCAGATTAAAGAGAAGCTGGTGACCCGCATCTACCGCGGGGTCGAGC 747
AAV_5_partial                   GAGTCAGATTCGCGCCCAGCTGGTGAAAGTGGTCTTCCAGGGAATTGAAC 611

AAV_IEE_from_serotype_2         --------------------------------------------------
AAV_6_complete                  CGACCCTGCCCAACTGGTTCGCGGTGACCAAGACGCGTAATGGCGCCGGA 745
```

Fig. 2 (continued)

```
AAV_1_complete                CGACCCTGCCCAACTGGTTCGCGGTGACCAAGACGCGTAATGGCGCCGGA 760
AAV_3_complete                CGCAGCTTCCGAACTGGTTCGCGGTGACCAAAACGCGAAATGGCGCCGGG 743
AAV_4_complete                CGCAGCTTCCGAACTGGTTCGCGGTGACCAAGACGCGTAATGGCGCCGGA 797
AAV_5_partial                 CCCAGATCAACGACTGGGTCGCCATCACCAAGGTAAAGAAGGGCG----- 656

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                GGGGGGAACAAGGTGGTGGACGAGTGCTACATCCCCAACTACCTCCTGCC 795
AAV_1_complete                GGGGGGAACAAGGTGGTGGACGAGTGCTACATCCCCAACTACCTCCTGCC 810
AAV_3_complete                GGCGGGAACAAGGTGGTGGACGACTGCTACATCCCCAACTACCTGCTCCC 793
AAV_4_complete                GGCGGGAACAAGGTGGTGGACGACTGCTACATCCCCAACTACCTGCTCCC 847
AAV_5_partial                 -GAGCCAATAAGGTGGTGGATTCTGGGTATATTCCCGCCTACCTGCTGCC 705

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                CAAGACTCAGCCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTATA 845
AAV_1_complete                CAAGACTCAGCCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTATA 860
AAV_3_complete                CAAGACCCAGCCCGAGCTCCAGTGGGCGTGGACTAACATGGACCAGTATT 843
AAV_4_complete                CAAGACCCAGCCCGAGCTCCAGTGGGCGTGGACTAACATGGACCAGTATA 897
AAV_5_partial                 GAAGGTCCAACCGGAGCTTCAGTGGGCGTGGACAAACCTGGACGAGTATA 755

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                TAAGCGCGTGTTTAAACCTGGCCGAGCGCAAACGGCTCGTGGCGCACGAC 895
AAV_1_complete                TAAGCGCCTGTTTGAACCTGGCCGAGCGCAAACGGCTCGTGGCGCAGCAC 910
AAV_3_complete                TAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACGGCTGGTGGCGCAGCAT 893
AAV_4_complete                TAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACGGCTGGTGGCGCAGCAT 947
AAV_5_partial                 AATTGGCCGCCCTGAATCTGGAGGAGCGCAAACGGCTCGTCGCGCAGTTT 805

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                CTGACCCACGTCAGCCAGACCCAGGAGCAGAACAAGGAGAATCTGAACCC 945
AAV_1_complete                CTGACCCACGTCAGCCAGACCCAGGAGCAGAACAAGGAGAATCTGAACCC 960
AAV_3_complete                CTGACGCACGTGTCGCAGACGCAGGAGCAGAACAAAGAGAATCAGAACCC 943
AAV_4_complete                CTGACGCACGTGTCGCAGACGCAGGAGCAGAACAAGGAAAACCAGAACCC 997
AAV_5_partial                 CTGGCAGAATCCTCGCAGCGCTCGCAGGAGGCGGCTTCGCAGCGTGAGTT 855

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                CAATTCTGACGCGCCTGTCATCCGGTCAAAAACCTCCGCACGCTACATGG 995
AAV_1_complete                CAATTCTGACGCGCCTGTCATCCGGTCAAAAACCTCCGCGCGCTACATGG 101
AAV_3_complete                CAATTCTGACGCGCCGGTCATCAGGTCAAAAACCTCAGCCAGGTACATGG 993
AAV_4_complete                CAATTCTGACGCGCCGGTCATCAGGTCAAAAACCTCCGCCAGGTACATGG 104
AAV_5_partial                 CTCGGCTGAC---CCGGTCATCAAAAGCAAGACTTCCCAGAAATACATGG 902

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                AGCTGGTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGG 104
AAV_1_complete                AGCTGGTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGG 106
AAV_3_complete                AGCTGGTCGGGTGGCTGGTGGACCGCGGGATCACGTCAGAAAAGCAATGG 104
AAV_4_complete                AGCTGGTCGGGTGGCTGGTGGACCGCGGGATCACGTCAGAAAAGCAATGG 109
AAV_5_partial                 CGCTCGTCAACTGGCTCGTGGAGCACGGCATCACTTCCGAGAAGCAGTGG 952

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                ATCCAGGAGGACCAGGCCTCGTACATCTCCTTCAACGCCGCCTCCAACTC 109
AAV_1_complete                ATCCAGGAGGACCAGGCCTCGTACATCTCCTTCAACGCCGCTTCCAACTC 111
AAV_3_complete                ATTCAGGAGGACCAGGCCTCGTACATCTCCTTCAACGCCGCCTCCAACTC 109
AAV_4_complete                ATCCAGGAGGACCAGGCGTCCTACATCTCCTTCAACGCCGCCTCCAACTC 114
AAV_5_partial                 ATCCAGGAAAATCAGGAGAGCTACCTCTCCTTCAACTCCACCGGCAACTC 100

AAV_IEE_from_serotype_2       --------------------------------------------------
```

Fig. 2 (continued)

```
AAV_6_complete                GCGGTCCCAGATCAAGGCCGCTCTGGACAATGCCGGCAAGATCATGGCGC 114
AAV_1_complete                GCGGTCCCAGATCAAGGCCGCTCTGGACAATGCCGGCAAGATCATGGCGC 116
AAV_3_complete                GCGGTCCCAGATCAAGGCCGCGCTGGACAATGCCTCCAAGATCATGAGCC 114
AAV_4_complete                GCGGTCACAAATCAAGGCCGCGCTGGACAATGCCTCCAAAATCATGAGCC 119
AAV_5_partial                 TCGGAGCCAGATCAAGGCCGCGCTCGACAACGCGACCAAAATTATGAGTC 105

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                TGACCAAATCCGCGCCCGACTACCTGGTAGGCCCCGCTCCGCCCGCCGAC 119
AAV_1_complete                TGACCAAATCCGCGCCCGACTACCTGGTAGGCCCCGCTCCGCCCGCGGAC 121
AAV_3_complete                TGACAAAGACGGCTCCGGACTACCTGGTGGGCAGCAACCCGCCGGAGGAC 119
AAV_4_complete                TGACAAAGACGGCTCCGGACTACCTGGTGGGCCAGAACCCGCCGGAGGAC 124
AAV_5_partial                 TGACAAAAAGCGCGGTGGACTACCTCGTGGGGAGCTCCGTTCCCGAGGAC 110

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                ATTAAAACCAACCGCATTTACCGCATCCTGGAGCTGAACGGCTACGACCC 124
AAV_1_complete                ATTAAAACCAACCGCATCTACCGCATCCTGGAGCTGAACGGCTACGAACC 126
AAV_3_complete                ATTACCAAAAATCGGATCTACCAAATCCTGGAGCTGAACGGGTACGATCC 124
AAV_4_complete                ATTTCCAGCAACCGCATCTACCGAATCCTCGAGATGAACGGGTACGATCC 129
AAV_5_partial                 ATTTCAAAAAACAGAATCTGGCAAATTTTTGAGATGAATGGCTACGACCC 115

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                TGCCTACGCCGGCTCCGTCTTTCTCGGCTGGGCCCAGAAAAGGTTCGGAA 129
AAV_1_complete                TGCCTACGCCGGCTCCGTCTTTCTCGGCTGGGCCCAGAAAAGGTTCGGGA 131
AAV_3_complete                GCAGTACGCGGCCTCCGTCTTCCTGGGCTGGGCGCAAAAGAAGTTCGGGA 129
AAV_4_complete                GCAGTACGCGGCCTCCGTCTTCCTGGGCTGGGCGCAAAAGAAGTTCGGGA 134
AAV_5_partial                 GGCCTACGCGGGATCCATCCTCTACGGCTGGTGTCAGCGCTCCTTCAACA 120

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                AACGCAACACCATCTGGCTGTTTGGGCCGGCCACCACGGGCAAGACCAAC 134
AAV_1_complete                AGCGCAACACCATCTGGCTGTTTGGGCCGGCCACCACGGGCAAGACCAAC 136
AAV_3_complete                AGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGACGGGTAAAACCAAC 134
AAV_4_complete                AGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGACGGGTAAAACCAAC 139
AAV_5_partial                 AGAGGAACACCGTCTGGCTCTACGGACCCGCCACGACCGGCAAGACCAAC 125

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                ATCGCGGAAGCCATCGCCCACGCCGTGCCCTTCTACGGCTGCGTCAACTG 139
AAV_1_complete                ATCGCGGAAGCCATCGCCCACGCCGTGCCCTTCTACGGCTGCGTCAACTG 141
AAV_3_complete                ATCGCGGAAGCCATCGCCCACGCCGTGCCCTTCTACGGCTGCGTAAACTG 139
AAV_4_complete                ATCGCGGAAGCCATCGCCCACGCCGTGCCCTTCTACGGCTGCGTGAACTG 144
AAV_5_partial                 ATCGCGGAGGCCATCGCCCACACTGTGCCCTTTTACGGCTGCGTGAACTG 130

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                GACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATCT 144
AAV_1_complete                GACCAATGAGAACTTTCCCTTCAATGATTGCGTCGACAAGATGGTGATCT 146
AAV_3_complete                GACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATCT 144
AAV_4_complete                GACCAATGAGAACTTTCCGTTCAACGATTGCGTCGACAAGATGGTGATCT 149
AAV_5_partial                 GACCAATGAAAACTTTCCCTTTAATGACTGTGTGGACAAAATGCTCATTT 135

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                GGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGTCCGCCAAGGCC 149
AAV_1_complete                GGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGTCCGCCAAGGCC 151
AAV_3_complete                GGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGAGCGCCAAGGCC 149
AAV_4_complete                GGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTAGAGAGCGCCAAGGCC 154
AAV_5_partial                 GGTGGGAGGAGGGAAAGATGACCAACAAGGTGGTTGAATCCGCCAAGGCC 140
```

Fig. 2 (continued)

```
AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              ATTCTCGGCGGCAGCAAGGTGCGCGTGGACCAAAAGTGCAAGTCGTCCGC 154
AAV_1_complete              ATTCTCGGCGGCAGCAAGGTGCGCGTGGACCAAAAGTGCAAGTCGTCCGC 156
AAV_3_complete              ATTCTGGGCGGAAGCAAGGTGCGCGTGGACCAAAAGTGCAAGTCATCGGC 154
AAV_4_complete              ATCCTGGGCGGAAGCAAGGTGCGCGTGGACCAAAAGTGCAAGTCATCGGC 159
AAV_5_partial               ATCCTGGGGGGCTCAAAGGTGCGGGTCGATCAGAAATGTAAATCCTCTGT 145

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              CCAGATCGATCCCACCCCCGTGATCGTCACCTCCAACACCAACATGTGCG 159
AAV_1_complete              CCAGATCGACCCCACCCCCGTGATCGTCACCTCCAACACCAACATGTGCG 161
AAV_3_complete              CCAGATCGAACCCACTCCCGTGATCGTCACCTCCAACACCAACATGTGCG 159
AAV_4_complete              CCAGATCGACCCAACTCCCGTGATCGTCACCTCCAACACCAACATGTGCG 164
AAV_5_partial               TCAAATTGATTCTACCCCTGTCATTGTAACTTCCAATACAAACATGTGTG 150

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              CCGTGATTGACGGGAACAGCACCACCTTCGAGCACCAGCAGCCGTTGCAG 164
AAV_1_complete              CCGTGATTGACGGGAACAGCACCACCTTCGAGCACCAGCAGCCGTTGCAG 166
AAV_3_complete              CCGTGATTGACGGGAACAGCACCACCTTCGAGCATCAGCAGCCGCTGCAG 164
AAV_4_complete              CGGTCATCGACGGAAACTCGACCACCTTCGAGCACCAACAACCACTCCAG 169
AAV_5_partial               TGGTGGTGGATGGGAATTCCACGACCTTTGAACACCAGCAGCCGCTGGAG 155

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGAGCATGACTTTGG 169
AAV_1_complete              GACCGGATGTTCAAATTTGAACTCACCCGCCGTCTGGAGCATGACTTTGG 171
AAV_3_complete              GACCGGATGTTTGAATTTGAACTTACCCGCCGTTTGGACCATGACTTTGG 169
AAV_4_complete              GACCGGATGTTCAAGTTCGAGCTCACCAAGCGCCTGGAGCACGACTTTGG 174
AAV_5_partial               GACCGCATGTTCAAATTTGAACTGACTAAGCGGCTCCCGCCAGATTTTGG 160

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              CAAGGTGACAAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCGCAGGATC 174
AAV_1_complete              CAAGGTGACAAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCGCAGGATC 176
AAV_3_complete              GAAGGTCACCAAACAGGAAGTAAAGGACTTTTTCCGGTGGGCTTCCGATC 174
AAV_4_complete              CAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCGTCAGATC 179
AAV_5_partial               CAAGATTACTAAGCAGGAAGTCAAGGACTTTTTTGCTTGGGCAAAGGTCA 165

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              ACGTGACCGAGGTGGCGCATGAGTTCTACGTCAGAAAGGGTGGAGCCAAC 179
AAV_1_complete              ACGTGACCGAGGTGGCGCATGAGTTCTACGTCAGAAAGGGTGGAGCCAAC 181
AAV_3_complete              ACGTGACTGACGTGGCTCATGAGTTCTACGTCAGAAAGGGTGGAGCTAAG 179
AAV_4_complete              ACGTGACCGAGGTGACTCACGAGTTTTACGTCAGAAAGGGTGGAGCTAGA 184
AAV_5_partial               ATCAGGTGCCGGTGACTCACGAGTTTAAAGTTCCCAGGGAATTGGCGGGA 170

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              AAGAGACCCGCCCCCGATGACGCGGATAAAAGCGAGCCCAAGCGGGCCTG 184
AAV_1_complete              AAAAGACCCGCCCCCGATGACGCGGATAAAAGCGAGCCCAAGCGGGCCTG 186
AAV_3_complete              AAACGCCCCGCCTCCAATGACGCGGATGTAAGCGAGCCAAAACGGGAGTG 184
AAV_4_complete              AAGAGGCCCGCCCCCAATGACGCAGATATAAGTGAGCCCAAGCGGGCCTG 189
AAV_5_partial               ACTAAAGGGGC-------GGAGAAATCTCTAAAACGCCCACTGGGTGACG 174

AAV_IEE_from_serotype_2     ----------------------------------------------------
AAV_6_complete              CCCCTCAGTCGCGGATCCATCGACGTCAGACGCGGAAGGAGCTCCGGTGG 189
AAV_1_complete              CCCCTCAGTCGCGGATCCATCGACGTCAGACGCGGAAGGAGCTCCGGTGG 191
AAV_3_complete              CACGTCACTTGCGCAGCCGACAACGTCAGACGCGGAAG---CACCGGCGG 189
AAV_4_complete              TCCGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAG---CTCCGGTGG 194
AAV_5_partial               TCACCAATACTAGCTATAAAAGTCTGGAGAAGCGGGCCAGGCTCTCATTT 179
```

Fig. 2 (continued)

```
AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           ACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTT 194
AAV_1_complete           ACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTT 196
AAV_3_complete           ACTACGCGGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGCATGAAT 194
AAV_4_complete           ACTACGCGGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGTATGAAT 199
AAV_5_partial            GTTCCCGAGACGCCTCGCAGTTCAGACGTGACTGTTGATCCCGCTC--CT 184

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           CAGATGCTGTTTCCCTGCAAAACATGCGAGAGAATGAATCAGAATTTCAA 199
AAV_1_complete           CAGATGCTGTTTCCCTGCAAGACATGCGAGAGAATGAATCAGAATTTCAA 201
AAV_3_complete           CTGATGCTTTTTCCCTGTAAAACATGCGAGAGAATGAATCAAATTTCCAA 199
AAV_4_complete           CTGATGCTTTTTCCCTGCCGGCAATGCGAGAGAATGAATCAGAATGTGGA 204
AAV_5_partial            CTGCGACCGCTCAATTGGAATTCAAGGTATGATTGCAAATGTGACTATCA 189

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           CATTTGCTTCACGCACGGGACCAGAGACTGTTCAGAATGTTTCCCCGGCG 204
AAV_1_complete           CATTTGCTTCACGCACGGGACGAGAGACTGTTCAGAGTGCTTCCCCGGCG 206
AAV_3_complete           TGTCTGTTTTACGCATGGTCAAAGAGACTGTGGGGAATGCTTCCCTGGAA 204
AAV_4_complete           CATTTGCTTCACGCACGGGGTCATGGACTGTGCCGAGTGCTTCCCCG--- 209
AAV_5_partial            TGCTCAATTTGACAACATTTCTAACAAATGTGATGAATG----------- 193

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           TGTCAGAATCTCAACCGGT------CGTCAGAAAGAGGACGTATCGGAAA 208
AAV_1_complete           TGTCAGAATCTCAACCGGT------CGTCAGAAAGAGGACGTATCGGAAA 210
AAV_3_complete           TGTCAGAATCTCAACCCGTTTCTGTCGTCAAAAAGAAGACTTATCAGAAA 209
AAV_4_complete           TGTCAGAATCTCAACCCGTGTCTGTCGTCAGAAAGCGGACGTATCAGAAA 214
AAV_5_partial            --TGAATATTTGAATCGGG-------GCAAAAATGGATGTATCTGTCACA 197

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           CTCTGTGCCATTCATCATCTGCTGGGGCGGGCTCCCGAGATTGCTTGCTC 213
AAV_1_complete           CTCTGTGCCATTCATCATCTGCTGGGGCGGGCTCCCGAGATTGCTTGCTC 215
AAV_3_complete           CTGTGTCCAATTCATCATATCCTGGGAAGGGCACCCGAGATTGCCTGTTC 214
AAV_4_complete           CTGTGTCCGATTCATCACATCATGGGGAGGGCGCCCGAGGTGGCCTGCTC 219
AAV_5_partial            ATGTAACTCACTGTCAAATTTGTCATGGGATTCCCCCCTGGGAAAAGGAA 202

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           GGCCTGCGATCTGGTCAACGTGGATCTGGATGACTGTGTTTCTGAGCAAT 218
AAV_1_complete           GGCCTGCGATCTGGTCAACGTGGACCTGGATGACTGTGTTTCTGAGCAAT 220
AAV_3_complete           GGCCTGCGATTTGGCCAATGTGGACTTGGATGACTGTGTTTCTGAGCAAT 219
AAV_4_complete           GGCCTGCGAACTGGCCAATGTGGACTTGGATGACTGTGACATGGAACAAT 224
AAV_5_partial            AACTTG----TCAGATTTTGGGGATTTTGACGATGCCAATAAAGAACAGT 206

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           AAATGACTTAAACCAGGTATGGCTGCCGATGGTTATCTTCCAGATTGGCT 223
AAV_1_complete           AAATGACTTAAACCAGGTATGGCTGCCGATGGTTATCTTCCAGATTGGCT 225
AAV_3_complete           AAATGACTTAAACCAGGTATGGCTGCTGACGGTTATCTTCCAGATTGGCT 224
AAV_4_complete           AAATGACTCAAACCAGATATGACT---GACGGTTACCTTCCAGATTGGCT 228
AAV_5_partial            AAATAAAGCGAGT-AGTCATGTCTTTTGTTGATCACCCTCCAGATTGGTT 211

AAV_IEE_from_serotype_2  --------------------------------------------------
AAV_6_complete           CGAGGACAACCTCTCTGAGGGCATTCGCGAGTGGTGGGACTTGAAACCTG 228
AAV_1_complete           CGAGGACAACCTCTCTGAGGGCATTCGCGAGTGGTGGGACTTGAAACCTG 230
AAV_3_complete           CGAGGACAACCTTTCTGAAGGCATTCGTGAGTGGTGGGCTCTGAAACCTG 229
AAV_4_complete           AGAGGACAACCTCTCTGAAGGCGTTCGAGAGTGGTGGGCGCTGCAACCTG 233
AAV_5_partial            GGAAGAAG---TTGGTGAAGGTCTTCGCGAGTTTTTGGGCCTTGAAGCGG 216
```

Fig. 2 (continued)

```
AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           GAGCCCCGAAACCCAAAGCCAACCAGCAAAAGCAGGACGACGGCCGGGGT 233
AAV_1_complete           GAGCCCCGAAGCCCAAAGCCAACCAGCAAAAGCAGGACGACGGCCGGGGT 235
AAV_3_complete           GAGTCCCTCAACCCAAAGCGAACCAACAACACCAGGACAACCGTCGGGGT 234
AAV_4_complete           GAGCCCCTAAACCCAAGGCAAATCAACAACATCAGGACAACGCTCGGGGT 238
AAV_5_partial            GCCCACCGAAACCAAAACCCAATCAGCAGCATCAAGATCAAGCCCGTGGT 221

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           CTGGTGCTTCCTGGCTACAAGTACCTCGGACCCTTCAACGGACTCGACAA 238
AAV_1_complete           CTGGTGCTTCCTGGCTACAAGTACCTCGGACCCTTCAACGGACTCGACAA 240
AAV_3_complete           CTTGTGCTTCCGGGTTACAAATACCTCGGACCCGGTAACGGACTCGACAA 239
AAV_4_complete           CTTGTGCTTCCGGGTTACAAATACCTCGGACCCGGCAACGGACTCGACAA 243
AAV_5_partial            CTTGTGCTGCCTGGTTATAACTATCTCGGACCCGGAAACGGTCTCGATCG 226

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           GGGGGAGCCCGTCAACGCGGCGGATGCAGCGGCCCTCGAGCACGACAAGG 243
AAV_1_complete           GGGGGAGCCCGTCAACGCGGCGGACGCAGCGGCCCTCGAGCACGACAAGG 245
AAV_3_complete           AGGAGAGCCGGTCAACGAGGCGGACGCGGCAGCCCTCGAACACGACAAAG 244
AAV_4_complete           GGGGGAACCCGTCAACGCAGCGGACGCGGCAGCCCTCGAGCACGACAAGG 248
AAV_5_partial            AGGAGAGCCTGTCAACAGGGCAGACGAGGTCGCGCGAGAGCACGACATCT 231

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           CCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAAC 248
AAV_1_complete           CCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAAC 250
AAV_3_complete           CTTACGACCAGCAGCTCAAGGCCGGTGACAACCCGTACCTCAAGTACAAC 249
AAV_4_complete           CCTACGACCAGCAGCTCAAGGCCGGTGACAACCCCTACCTCAAGTACAAC 253
AAV_5_partial            CGTACAACGAGCAGCTTGAGGCGGGAGACAACCCCTACCTCAAGTACAAC 236

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           CACGCCGACGCCGAGTTTCAGGAGCGTCTGCAAGAAGATACGTCTTTTGG 253
AAV_1_complete           CACGCCGACGCCGAGTTTCAGGAGCGTCTGCAAGAAGATACGTCTTTTGG 255
AAV_3_complete           CACGCCGACGCCGAGTTTCAGGAGCGTCTTCAAGAAGATACGTCTTTTGG 254
AAV_4_complete           CACGCCGACGCGGAGTTCCAGCAGCGGCTTCAGGGCGACACATCGTTTGG 258
AAV_5_partial            CACGCGGACGCCGAGTTTCAGGAGAAGCTCGCCGACGACACATCCTTCGG 241

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           GGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGAGGGTTCTCGAAC 258
AAV_1_complete           GGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGCGGGTTCTCGAAC 260
AAV_3_complete           GGGCAACCTTGGCAGAGCAGTCTTCCAGGCCAAAAAGAGGATCCTTGAGC 259
AAV_4_complete           GGGCAACCTCGGCAGAGCAGTCTTCCAGGCCAAAAAGAGGGTTCTTGAAC 263
AAV_5_partial            GGGAAACCTCGGAAAGGCAGTCTTTCAGGCCAAGAAAAGGGTTCTCGAAC 246

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           CTTTTGGTCTGGTTGAGGAAGGTGCTAAGACGGCTCCTGGAAAGAAACGT 263
AAV_1_complete           CTCTCGGTCTGGTTGAGGAAGGCGCTAAGACGGCTCCTGGAAAGAAACGT 265
AAV_3_complete           CTCTTGGTCTGGTTGAGGAAGCAGCTAAAACGGCTCCTGGAAAGAAGGGG 264
AAV_4_complete           CTCTTGGTCTGGTTGAGCAAGCGGGTGAGACGGCTCCTGGAAAGAAGAGA 268
AAV_5_partial            CTTTTGGCCTGGTTGAAGAGGGTGCTAAGACGGCCCCTACCGGAAAGCGG 251

AAV_IEE_from_serotype_2  ----------------------------------------------------
AAV_6_complete           CCGGTAGAGCAGTCGCCACAAGAGCCAGACTCCTCCTCGGGCATTGGCAA 268
AAV_1_complete           CCGGTAGAGCAGTCGCCACAAGAGCCAGACTCCTCCTCGGGCATCGGCAA 270
AAV_3_complete           GCTGTAGATCAGTCTCCTCAGGAACCGGACTCATCATCTGGTGTTGGCAA 269
AAV_4_complete           CCGTTGATTGAATCCCCCCAGCAGCCCGACTCCTCCACGGGTATCGGCAA 273
```

Fig. 2 (continued)

```
AAV_5_partial                ATAGACGACCACTTTCCAAAAAGAAAGAAGGCTCGGAC---------CGA 255

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               GACAGGCCAGCAGCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGGCG 273
AAV_1_complete               GACAGGCCAGCAGCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGGCG 275
AAV_3_complete               ATCGGGCAAACAGCCTGCCAGAAAAAGACTAAATTTCGGTCAGACTGGAG 274
AAV_4_complete               AAAAGGCAAGCAGCCGGCTAAAAAGAAGCTCGTTTTCGAAGACG-----A 278
AAV_5_partial                AGAGGACTCCAAGCCTTCCA--------------CCTCGTCAGACGCCGA 259

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               ACTCAGAGTCAGTCCCCGACCCACAACCTCTCGGAGAACCTCCAGCAACC 278
AAV_1_complete               ACTCAGAGTCAGTCCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACC 280
AAV_3_complete               ACTCAGAGTCAGTCCCAGACCCTCAACCTCTCGGAGAACCACCAGCAGCC 279
AAV_4_complete               AACTGGAGCAGGCGACGGACCC----CCTGAGGGATCAACTTCCGGAGCC 282
AAV_5_partial                AGCTGGACCCAG---CGGATCC-CAGCAGCTGCAAATCCCAGCCCAACCA 263

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               CCCGCTGCTGTGGGACCTACTACAATGGCTTCAGGCGGTGGCGCACCAAT 283
AAV_1_complete               CCCGCTGCTGTGGGACCTACTACAATGGCTTCAGGCGGTGGCGCACCAAT 285
AAV_3_complete               CCCACAAGTTTGGGATCTAATACAATGGCTTCAGGCGGTGGCGCACCAAT 284
AAV_4_complete               ATGTCTGATGA------CAGTGAGATGCGTGCAGCAGCTGGCGGAGCTGC 287
AAV_5_partial                GCCTCAAGTTTGGGAGCTGATACAATGTCTGCGGGAGGTGGCGGCCCATT 268

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               GGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAATGCCTCAGGAAATT 288
AAV_1_complete               GGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAATGCCTCAGGAAATT 290
AAV_3_complete               GGCAGACAATAACGAGGGTGCCGATGGAGTGGGTAATTCCTCAGGAAATT 289
AAV_4_complete               AGTCGAGGGCGGACAAGGTGCCGATGGAGTGGGTAATGCCTCGGGTGATT 292
AAV_5_partial                GGGCGACAATAACCAAGGTGCCGATGGAGTGGGCAATGCCTCGGGAGATT 273

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               GGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACC 293
AAV_1_complete               GGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACC 295
AAV_3_complete               GGCATTGCGATTCCCAATGGCTGGGCGACAGAGTCATCACCACCAGCACC 294
AAV_4_complete               GGCATTGCGATTCCACCTGGTCTGAGGGCCACGTCACGACCACCAGCACC 297
AAV_5_partial                GGCATTGCGATTCCACGTGGATGGGGGACAGAGTCGTCACCAAGTCCACC 278

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               CGAACATGGGCCTTGCCCACCTATAACAACCACCTCTACAAGCAAATCTC 298
AAV_1_complete               CGCACCTGGGCCTTGCCCACCTACAATAACCACCTCTACAAGCAAATCTC 300
AAV_3_complete               AGAACCTGGGCCCTGCCCACTTACAACAACCATCTCTACAAGCAAATCTC 299
AAV_4_complete               AGAACCTGGGTCTTGCCCACCTACAACAACCACCTCTACAAGCGACTC-- 302
AAV_5_partial                CGAACCTGGGTGCTGCCCAGCTACAACAACCACCAGTACCGAGAGATCAA 283

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               CAGTGCTTCAACGGGGGCCAGCAACGACAACCACTACTTCGGCTACAGCA 303
AAV_1_complete               CAGTGCTTCAACGGGGGCCAGCAACGACAACCACTACTTCGGCTACAGCA 305
AAV_3_complete               CAGC---CAATCAGGAGCTTCAAACGACAACCACTACTTTGGCTACAGCA 303
AAV_4_complete               ---------GGAGAGAGCCTGCAGTC-CAACACCTACAACGGATTCTCCA 306
AAV_5_partial                AAGCGGCTCCGTCGACGGAAGCAACGCCAACGCCTACTTTGGATACAGCA 288

AAV_IEE_from_serotype_2      --------------------------------------------------
AAV_6_complete               CCCCCTGGGGGTATTTTGATTTCAACAGATTCCACTGCCATTTCTCACCA 308
AAV_1_complete               CCCCCTGGGGGTATTTTGATTTCAACAGATTCCACTGCCACTTTTCACCA 310
AAV_3_complete               CCCCTTGGGGGTATTTTGACTTTAACAGATTCCACTGCCACTTCTCACCA 308
```

Fig. 2 (continued)

```
AAV_4_complete              CCCCCTGGGGATACTTTGACTTCAACCGCTTCCACTGCCACTTCTCACCA 311
AAV_5_partial               CCCCCTGGGGGTACTTTGACTTTAACCGCTTCCACAGCCACTGGAGCCCC 293

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              CGTGACTGGCAGCGACTCATCAACAACAATTGGGGATTCCGGCCCAAGAG 313
AAV_1_complete              CGTGACTGGCAGCGACTCATCAACAACAATTGGGGATTCCGGCCCAAGAG 315
AAV_3_complete              CGTGACTGGCAGCGACTCATTAACAACAACTGGGGATTCCGGCCCAAGAA 313
AAV_4_complete              CGTGACTGGCAGCGACTCATCAACAACAACTGGGGCATGCGACCCAAAGC 316
AAV_5_partial               CGAGACTGGCAAAGACTCATCAACAACTACTGGGGCTTCAGACCCCGGTC 298

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              ACTCAACTTCAAGCTCTTCAACATCCAAGTCAAGGAGGTCACGACGAATG 318
AAV_1_complete              ACTCAACTTCAAACTCTTCAACATCCAAGTCAAGGAGGTCACGACGAATG 320
AAV_3_complete              ACTCAGCTTCAAGCTCTTCAACATCCAAGTTAGAGGGGTCACGCAGAACG 318
AAV_4_complete              CATGCGGGTCAAAATCTTCAACATCCAGGTCAAGGAGGTCACGACGTCGA 321
AAV_5_partial               CCTCAGAGTCAAAATCTTCAACATTCAAGTCAAAGAGGTCACGGTGCAGG 303

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              ATGGCGTCACGACCATCGCTAATAACCTTACCAGCACGGTTCAAGTCTTC 323
AAV_1_complete              ATGGCGTCACAACCATCGCTAATAACCTTACCAGCACGGTTCAAGTCTTC 325
AAV_3_complete              ATGGCACGACGACTATTGCCAATAACCTTACCAGCACGGTTCAAGTGTTT 323
AAV_4_complete              ACGGCGAGACAACGGTGGCTAATAACCTTACCAGCACGGTTCAGATCTTT 326
AAV_5_partial               ACTCCACCACCACCATCGCCAACAACCTCACCTCCACCGTCCAAGTGTTT 308

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              TCGGACTCGGAGTACCAGTTGCCGTACGTCCTCGGCTCTGCGCACCAGGG 328
AAV_1_complete              TCGGACTCGGAGTACCAGCTTCCGTACGTCCTCGGCTCTGCGCACCAGGG 330
AAV_3_complete              ACGGACTCGGAGTATCAGCTCCCGTACGTGCTCGGGTCGGCGCACCAAGG 328
AAV_4_complete              GCGGACTCGTCGTACGAACTGCCGTACGTGATGGATGCGGGTCAAGAGGG 331
AAV_5_partial               ACGGACGACGACTACCAGCTGCCCTACGTCGTCGGCAACGGGACCGAGGG 313

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              CTGCCTCCCTCCGTTCCCGGCGGACGTGTTCATGATTCCGCAGTACGGCT 333
AAV_1_complete              CTGCCTCCCTCCGTTCCCGGCGGACGTGTTCATGATTCCGCAATACGGCT 335
AAV_3_complete              CTGTCTCCCGCCGTTTCCAGCGGACGTCTTCATGGTCCCTCAGTATGGAT 333
AAV_4_complete              CAGCCTGCCTCCTTTTCCCAACGACGTCTTTATGGTGCCCCAGTACGGCT 336
AAV_5_partial               ATGCCTGCCGGCCTTCCCTCCGCAGGTCTTTACGCTGCCGCAGTACGGTT 318

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              ACCTAACGCTCAACAATGGCAGC---------CAGGCAGTGGGACGGTCA 338
AAV_1_complete              ACCTGACGCTCAACAATGGCAGC---------CAAGCCGTGGGACGTTCA 339
AAV_3_complete              ACCTCACCCTGAACAACGGAAGT---------CAAGCGGTGGGACGCTCA 337
AAV_4_complete              ACTGTGGACTGGTGACCGGCAACACTTCGCAGCAACAGACTGACAGAAAT 341
AAV_5_partial               ACGCGACGCTGAACCGCGACAACAC---AGAAAATCCCACCGAGAGGAGC 323

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              TCCTTTTACTGCCTGGAATATTTCCCATCGCAGATGCTGAGAACGGGCAA 343
AAV_1_complete              TCCTTTTACTGCCTGGAATATTTCCCTTCTCAGATGCTGAGAACGGGCAA 344
AAV_3_complete              TCCTTTTACTGCCTGGAGTACTTCCCTTCGCAGATGCTAAGGACTGGAAA 342
AAV_4_complete              GCCTTCTACTGCCTGGAGTACTTTCCTTCGCAGATGCTGCGGACTGGCAA 346
AAV_5_partial               AGCTTCTTCTGCCTAGAGTACTTTCCCAGCAAGATGCTGAGAACGGGCAA 328

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              TAACTTTACCTTCAGCTACACCTTCGAGGACGTGCCTTTCCACAGCAGCT 348
AAV_1_complete              CAACTTTACCTTCAGCTACACCTTTGAGGAAGTGCCTTTCCACAGCAGCT 349
```

Fig. 2 (continued)

```
AAV_3_complete              TAACTTCCAATTCAGCTATACCTTCGAGGATGTACCTTTTCACAGCAGCT 347
AAV_4_complete              CAACTTTGAAATTACGTACAGTTTTGAGAAGGTGCCTTTCCACTCGATGT 351
AAV_5_partial               CAACTTTGAGTTTACCTACAACTTTGAGGAGGTGCCCTTCCACTCCAGCT 333

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              ACGCGCACAGCCAGAGCCTGGACCGGCTGATGAATCCTCTCATCGACCAG 353
AAV_1_complete              ACGCGCACAGCCAGAGCCTGGACCGGCTGATGAATCCTCTCATCGACCAA 354
AAV_3_complete              ACGCTCACAGCCAGAGTTTGGATCGCTTGATGAATCCTCTTATTGATCAG 352
AAV_4_complete              ACGCGCACAGCCAGAGCCTGGACCGGCTGATGAACCCTCTCATCGACCAG 356
AAV_5_partial               TCGCTCCCAGTCAGAACCTCTTCAAGCTGGCCAACCCGCTGGTGGACCAG 338

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              TACCTGTATTACCTGAACAGAACTCAGAATCAG---TCCGGAAGTGCCCA 357
AAV_1_complete              TACCTGTATTACCTGAACAGAACTCAAAATCAG---TCCGGAAGTGCCCA 359
AAV_3_complete              TATCTGTACTACCTGAACAGAACGCAAGGAACAACCTCTGGAACAACCAA 357
AAV_4_complete              TACCTGTGGGGACTGCAATCGACCACCACCGGAACCACCCTGAATGCCGG 361
AAV_5_partial               TACTTGTACCGCTTCGTGAGCACAAATA--------------------A 341

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              AAACAAGGACTTGCTGTTTAGCCGGGGGTCTCCAGCTGGCATGTCTGTTC 362
AAV_1_complete              AAACAAGGACTTGCTGTTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTC 364
AAV_3_complete              CCAATCACGGCTGCTTTTTAGCCAGGCTGGGCCTCAGTCTATGTCTTTGC 362
AAV_4_complete              GACTGCCACCACCAACTTTACCAAGCTGCGGCCTACCAACTTTTCCAACT 366
AAV_5_partial               CACTGGCGGAGTCCAGTTCAACAAGAACCTGGCCGGGAGATACGCCAACA 346

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              AGCCCAAAAACTGGCTACCTGGACCCTGTTACCGGCAGCAGCGCGTTTCT 367
AAV_1_complete              AGCCCAAAAACTGGCTACCTGGACCCTGTTATCGGCAGCAGCGCGTTTCT 369
AAV_3_complete              AGGCCAGAAATTGGCTACCTGGGCCCTGCTACCGGCAACAGAGACTTTCA 367
AAV_4_complete              TTAAAAAGAACTGGCTGCCCGGGCCTTCAATCAAGCAGCAGGGCTTCTCA 371
AAV_5_partial               CCTACAAAAACTGGTTCCCGGGGCCCATGGGCCGAACCCAGGGCTGGAAC 351

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              AAAACAAAAACAGACAACAACAACAGCAACTTTACCTGGACTGGTG---C 372
AAV_1_complete              AAAACAAAAACAGACAACAACAACAGCAATTTTACCTGGACTGGTG---C 373
AAV_3_complete              AAGACTGCTAACGACAACAACAACAGTAACTTTCCTTGGACAGCGG---C 372
AAV_4_complete              AAGACTGCCAATCAAAACTACAAGATCCCTGCCACCGGGTCAGACAGTCT 376
AAV_5_partial               CTGGGCTCCGGGGTCAACCGCGCCAGTGTCAGCGCCTTCGCCACGA--CC 356

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              TTCAAAATATAA------------CCTTAATGGGCGTGAATCTATAATCA 376
AAV_1_complete              TTCAAAATATAA------------CCTCAATGGGCGTGAATCCATCATCA 377
AAV_3_complete              CAGCAAATATCA------------TCTCAATGGCCGCGACTCGCTGGTGA 376
AAV_4_complete              CATCAAATACGAGACGCACAGCACTCTGGACGGAAGATGGAGTGCCCTGA 381
AAV_5_partial               AATAGGATGGAGCTCGAGGGCGCGAGTTACCAGGTGCCCCCGCAGCCGAA 361

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              ACCCTGGCACTGCTATGGCCTCACACAAAGACGACAAAGACAAGTTCTTT 381
AAV_1_complete              ACCCTGGCACTGCTATGGCCTCACACAAAGACGACGAAGACAAGTTCTTT 382
AAV_3_complete              ATCCAGGACCAGCTATGGCCAGTCACAAGGACGATGAAGAAAAATTTTTC 381
AAV_4_complete              CCCCCGGACCTCCAATGGCCACGGCTGGACCTGCGGACAGCAAGTTCAGC 386
AAV_5_partial               CGGCATGACCAACAACCTCCAGGG-CAGCAACACCTATGCCCTGGAGAAC 366

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              CCCATGAGCGGTGTCATGATTTTTGGAAAGGAGAGCGCCGGAGCTTCAAA 386
```

Fig. 2 (continued)

```
AAV_1_complete              CCCATGAGCGGTGTCATGATTTTTGGAAAAGAGAGCGCCGGAGCTTCAAA 387
AAV_3_complete              CCTATGCACGGCAATCTAATATTTGGCAAAGAAGGGACAACGGCAAGTAA 386
AAV_4_complete              AACAGCCAGCTCATCTT--TGCGGGGCCTAAACAGAACGGCAACACG-GC 390
AAV_5_partial               ACTATGATCTTCAACAGCCAGCCGGCGAACCCGGGCACCACCGCCACGTA 371

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              CACTGCATTGGACAATGTCATGATCACAGACGAAGAGGAAATCAAAGCCA 391
AAV_1_complete              CACTGCATTGGACAATGTCATGATTACAGACGAAGAGGAAATTAAAGCCA 392
AAV_3_complete              CGCAGAATTAGATAATGTAATGATTACGGATGAAGAAGAGATTCGTACCA 391
AAV_4_complete              CACCGTACCCGGGACTCTGATCTTCACCTCTGAGGAGGAGCTGGCAGCCA 395
AAV_5_partial               CCTCGAG---GGCAACATGCTCATCACCAGCGAGAGCGAGACGCAGCCGG 375

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              CTAACCCCGTGGCCACCGAAAGATTTGGGACTGTGGCAGTCAATCTCCAG 396
AAV_1_complete              CTAACCCTGTGGCCACCGAAAGATTTGGGACCGTGGCAGTCAATTTCCAG 397
AAV_3_complete              CCAATCCTGTGGCAACAGAGCAGTATGGAACTGTGGCAAATAACTTGCAG 396
AAV_4_complete              CCAACGCCACCGATACGGACATGTGGGGCAACCTACCTGGCGGTGACCAG 400
AAV_5_partial               TGAACCGCGTGGCGTACAACGTCGGCGGGCAGATGGCCACCAACAACCAG 380

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              AGCAGCAGCACAGACCCTGCGACCGGAGATGTGCATGTTATGGGAGCCTT 401
AAV_1_complete              AGCAGCAGCACAGACCCTGCGACCGGAGATGTGCATGCTATGGGAGCATT 402
AAV_3_complete              AGCTCAAATACAGCTCCCACGACTGGAACTGTCAATCATCAGGGGGCCTT 401
AAV_4_complete              AGCAACAGCAACCTGCCGACCGTGGACAGACTGACAGCCTTGGGAGCCGT 405
AAV_5_partial               AGCTCCACCACTGCCCCCGCGACCGGCACGTACAACCTCCAGGAAATCGT 385

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              ACCTGGAATGGTGTGGCAAGACAGAGACGTATACCTGCAGGGTCCTATTT 406
AAV_1_complete              ACCTGGCATGGTGTGGCAAGATAGAGACGTGTACCTGCAGGGTCCCATTT 407
AAV_3_complete              ACCTGGCATGGTGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCT 406
AAV_4_complete              GCCTGGAATGGTCTGGCAAAACAGAGACATTTACTACCAGGGTCCCATTT 410
AAV_5_partial               GCCCGGCAGCGTGTGGATGGAGAGGGACGTGTACCTCCAAGGACCCATCT 390

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              GGGCCAAAATTCCTCACACGGATGGACACTTTCACCCGTCTCCTCTCATG 411
AAV_1_complete              GGGCCAAAATTCCTCACACAGATGGACACTTTCACCCGTCTCCTCTTATG 412
AAV_3_complete              GGGCAAAGATTCCTCACACGGATGGACACTTTCATCCTTCTCCTCTGATG 411
AAV_4_complete              GGGCCAAGATTCCTCATACCGATGGACACTTTCACCCCTCACCGCTGATT 415
AAV_5_partial               GGGCCAAGATCCCAGAGACGGGGGCGCACTTTCACCCCTCTCCGGCCATG 395

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              GGCGGCTTTGGACTTAAGCACCCGCCTCCTCAGATCCTCATCAAAAACAC 416
AAV_1_complete              GGCGGCTTTGGACTCAAGAACCCGCCTCCTCAGATCCTCATCAAAAACAC 417
AAV_3_complete              GGAGGCTTTGGACTGAAACATCCGCCTCCTCAAATCATGATCAAAAATAC 416
AAV_4_complete              GGTGGGTTTGGGCTGAAACACCCGCCTCCTCAAATTTTTATCAAGAACAC 420
AAV_5_partial               GGCGGATTCGGACTCAAACACCCACCGCCCATGATGCTCATCAAGAACAC 400

AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              GCCTGTTCCTGCGAATCCTCCGGCAGAGTTTTCGGCTACAAAGTTTGCTT 421
AAV_1_complete              GCCTGTTCCTGCGAATCCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTT 422
AAV_3_complete              TCCGGTACCGGCAAATCCTCCGACGACTTTCAGCCCGGCCAAGTTTGCTT 421
AAV_4_complete              CCCGGTACCTGCGAATCCTGCAACGACCTTCAGCTCTACTCCGGTAAACT 425
AAV_5_partial               GCCTGTGCCCGGAAATATC---ACCAGCTTCTCGGACGTGCCCGTCAGCA 405

AAV_IEE_from_serotype_2     --------------------------------------------------
```

Fig. 2 (continued)

```
AAV_6_complete                CATTCATCACCCAGTATTCCACAGGACAAGTGAGCGTGGAGATTGAATGG 426
AAV_1_complete                CATTCATCACCCAATACTCCACAGGACAAGTGAGTGTGGAAATTGAATGG 427
AAV_3_complete                CATTTATCACTCAGTACTCCACTGGACAGGTCAGCGTGGAAATTGAGTGG 426
AAV_4_complete                CCTTCATTACTCAGTACAGCACTGGCCAGGTGTCGGTGCAGATTGACTGG 430
AAV_5_partial                 GCTTCATCACCCAGTACAGCACCGGGCAGGTCACCGTGGAGATGGAGTGG 410

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                GAGCTGCAGAAAGAAAACAGCAAACGCTGGAATCCCGAAGTGCAGTATAC 431
AAV_1_complete                GAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCCGAAGTGCAGTACAC 432
AAV_3_complete                GAGCTACAGAAAGAAAACAGCAAACGTTGGAATCCAGAGATTCAGTACAC 431
AAV_4_complete                GAGATCCAGAAGGAGCGGTCCAAACGCTGGAACCCCGAGGTCCAGTTTAC 435
AAV_5_partial                 GAGCTCAAGAAGGAAAACTCCAAGAGGTGGAACCCAGAGATCCAGTACAC 415

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                ATCTAACTATGCAAAATCTGCCAACGTTGATTTCACTGTGGACAACAATG 436
AAV_1_complete                ATCCAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAACAATG 437
AAV_3_complete                TTCCAACTACAACAAGTCTGTTAATGTGGACTTTACTGTAGACACTAATG 436
AAV_4_complete                CTCCAACTACGGACAGCAAAACTCTCTGTTGTGGGCTCCCGATGCGGCTG 440
AAV_5_partial                 AAACAACTACAACGACCCCCAGTTTGTGGACTTTGCCCCGGACAGCACCG 420

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                GACTTTATACTGAGCCTCGCCCCATTGGCACCCGTTACCTCACCCGTCCC 441
AAV_1_complete                GACTTTATACTGAGCCTCGCCCCATTGGCACCCGTTACCTTACCCGTCCC 442
AAV_3_complete                GTGTTTATAGTGAACCTCGCCCTATTGGAACCCGGTATCTCACACGAAAC 441
AAV_4_complete                GGAAATACACTGAGCCTAGGGCTATCGGTACCCGCTACCTCACCCACCAC 445
AAV_5_partial                 GGGAATACAGAACCACCAGACCTATCGGAACCCGATACCTTACCCGACCC 425

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                CTGTAATTGTGTGTTAA---------TCAATAAACCGGTTAATTCGTGTC 445
AAV_1_complete                CTGTAATTACGTGTTAA---------TCAATAAACCGGTTGATTCGTTTC 446
AAV_3_complete                TTGTGAATCCTGGTTAA---------TCAATAAACCGTTTAATTCGTTTC 445
AAV_4_complete                CTGTAATAACCTGTTAA---------TCAATAAACCGGTTTATTCGTTTC 449
AAV_5_partial                 CTTTAACCCATTCATGTCGCATACCCTCAATAAACCG-TGTATTCGTGTC 430

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                AGTTGAACTTTGGTCTCA---TGTCGTTATTATCTTATCTGGTCACCATA 450
AAV_1_complete                AGTTGAACTTTGGTCTCC---TGTCCTTCTTATCTTATC-GGTTACCATG 451
AAV_3_complete                AGTTGAACTTTGGCTCTTGTGCACTTCTTTATCTTTATCTTGTTTCCATG 450
AAV_4_complete                AGTTGAACTTTGGTCTCCG--TGTCCTTCTTATCTTATCTCGTTTCCATG 454
AAV_5_partial                 AGTAAAATACTGCCTCTTGT-GGTCATTCAATGAATAACAGCTTACAACA 435

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                GCAACCGGTTACACATTAACTGCTTAG------TTGCGCTTCGCGAATAC 454
AAV_1_complete                GTTATAGCTTACACATTAACTGCTTGG------TTGCGCTTCGCGATAAA 455
AAV_3_complete                GCTACTGCGTAGATAAGCAGCGGCCTGCGGCGCTTGCGCTTCGCGGTTTA 455
AAV_4_complete                GCTACTGCGTACATAAGCAGCGGCCTGCGGCGCTTGCGCTTCGCGGTTTA 459
AAV_5_partial                 TCTACAAAACCTCCTTGCTTGAGAGTGTGGCACTCTCCCCCC-TGTCGCG 440

AAV_IEE_from_serotype_2       --------------------------------------------------
AAV_6_complete                CCCTA------------GTGATGGAGTTGCCCACTCCCTCTATGCGCGCT 458
AAV_1_complete                AGACT------------TACGTCATCGGGTTACCCCTAGTGATGGAGTTG 459
AAV_3_complete                CAACTGCTGGTTAATATTTAACTCTCGCCATACCTCTAGTGATGGAGTTG 460
AAV_4_complete                CAACTGCCGGTTAATCAGTAACTTCTGGCAAACC--AGATGATGGAGTTG 464
AAV_5_partial                 CG---------------------------------------------- 440
```

Fig. 2 (continued)

```
AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              CGCTCGCTCGGTGGGGCCGGCAGAGCAGAGCTCTGCCGTCTGCGGACCTT 463
AAV_1_complete              CCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCCTGCGGACCAAA 464
AAV_3_complete              GCCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGGCCTGGCGACCAAA 465
AAV_4_complete              GCCACATTAGCTATGCGCGCTCGCTCACTCACTCGGCCCTGGAGACCAAA 469
AAV_5_partial               --------------------------------------------------


AAV_IEE_from_serotype_2     --------------------------------------------------
AAV_6_complete              TGGTCCGCAGGCCCCACCGAGCGAGCGAGCGCGCATAGAGGGAGTGGGCA 468
AAV_1_complete              GGTCCGCAGACGGCAGAGCTCTGCTCTGCCGGCCCCACCGAGCGAGCGAG 469
AAV_3_complete              GGTCGCCAGACGGACGTGCTTTGCACGTCCGGCCCCACCGAGCGAGCGAG 470
AAV_4_complete              GGTCTCCAGACTGCCGGCCTCTGGCCGGCAGGGCCGAGTGAGTGAGCGAG 474
AAV_5_partial               --------------------------------------------------


AAV_IEE_from_serotype_2     ----------------------
AAV_6_complete              A---------------------- 4683
AAV_1_complete              CGCGCAGAGAGGGAGTGGGCAA 4718
AAV_3_complete              TGCGCATAGAGGGAGTGGCCAA 4726
AAV_4_complete              CGCGCATAGAGGGAGTGGCCAA 4767
AAV_5_partial               ----------------------
```

Fig. 2 (continued)

221916.ST25
SEQUENCE LISTING

<110> CORNELL RESEARCH FOUNDATION, INC.
FALCK-PEDERSEN, ERIK S
PHILPOTT, NICOLA

<120> USE OF AAV INTEGRATION EFFICIENCY ELEMENT FOR MEDIATING SITE-SPECIFIC INTEGRATION OF A TRANSCRIPTION UNIT

<130> 221916

<150> US 60/371,044
<151> 2002-04-09

<160> 6

<170> PatentIn version 3.2

<210> 1
<211> 102
<212> DNA
<213> adeno-associated virus 2

<400> 1
cgacattttg cgacaccatg tggtcacgct gggtatttaa gcccgagtga gcacgcaggg 60

tctccatttt gaagcgggag gtttgaacgc gcagccgcca tg 102

<210> 2
<211> 4683
<212> DNA
<213> adeno-associated virus serotype 6

<400> 2
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc 60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgcag agagggagtg 120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag 180

ggttagggag gtcctgtatt agaggtcacg tgagtgtttt gcgacatttt gcgacaccat 240

gtggtcacgc tgggtattta agcccgagtg agcacgcagg gtctccattt tgaagcggga 300

ggtttgaacg cgcagcgcca tgccggggtt ttacgagatt gtgattaagg tccccagcga 360

ccttgacgag catctgcccg gcatttctga cagctttgtg aactgggtgg ccgagaagga 420

atgggagttg ccgccagatt ctgacatgga tctgaatctg attgagcagg cacccctgac 480

cgtggccgag aagctgcagc gcgacttcct ggtccagtgg cgccgcgtga gtaaggcccc 540

ggaggccctc ttctttgttc agttcgagaa gggcgagtcc tacttccacc tccatattct 600

ggtggagacc acgggggtca aatccatggt gctgggccgc ttcctgagtc agattaggga 660

caagctggtg cagaccatct accgcgggat cgagccgacc ctgcccaact ggttcgcggt 720

gaccaagacg cgtaatggcg ccggaggggg gaacaaggtg gtggacgagt gctacatccc 780

caactacctc ctgcccaaga ctcagcccga gctgcagtgg gcgtggacta acatggagga 840

gtatataagc gcgtgtttaa acctggccga gcgcaaacgg ctcgtggcgc acgacctgac 900

ccacgtcagc cagacccagg agcagaacaa ggagaatctg aaccccaatt ctgacgcgcc 960

tgtcatccgg tcaaaaacct ccgcacgcta catggagctg gtcgggtggc tggtggaccg 1020

gggcatcacc tccgagaagc agtggatcca ggaggaccag gcctcgtaca tctccttcaa 1080

cgccgcctcc aactcgcggt cccagatcaa ggccgctctg gacaatgccg gcaagatcat 1140

ggcgctgacc aaatccgcgc ccgactacct ggtaggcccc gctccgcccg ccgacattaa 1200

aaccaaccgc atttaccgca tcctggagct gaacggctac gaccctgcct acgccggctc 1260

cgtctttctc ggctgggccc agaaaaggtt cggaaaacgc aacaccatct ggctgtttgg 1320

gccggccacc acgggcaaga ccaacatcgc ggaagccatc gcccacgccg tgcccttcta 1380

cggctgcgtc aactggacca atgagaactt tcccttcaac gattgcgtcg acaagatggt 1440

gatctggtgg gaggagggca agatgacggc caaggtcgtg gagtccgcca aggccattct 1500

cggcggcagc aaggtgcgcg tggaccaaaa gtgcaagtcg tccgcccaga tcgatcccac 1560

ccccgtgatc gtcacctcca acaccaacat gtgcgccgtg attgacggga acagcaccac 1620

cttcgagcac cagcagccgt tgcaggaccg gatgttcaaa tttgaactca cccgccgtct 1680

ggagcatgac tttggcaagg tgacaaagca ggaagtcaaa gagttcttcc gctgggcgca 1740

ggatcacgtg accgaggtgg cgcatgagtt ctacgtcaga aagggtggag ccaacaagag 1800

acccgccccc gatgacgcgg ataaaagcga gcccaagcgg gcctgccgct cagtcgcgga 1860

tccatcgacg tcagacgcgg aaggagctcc ggtggacttt gccgacaggt accaaaacaa 1920

atgttctcgt cacgcgggca tgcttcagat gctgtttccc tgcaaaacat gcgagagaat 1980

gaatcagaat ttcaacattt gcttcacgca cgggaccaga gactgttcag aatgtttccc 2040

cggcgtgtca gaatctcaac cggtcgtcag aaagaggacg tatcggaaac tctgtgccat 2100

tcatcatctg ctggggcggg ctcccgagat tgcttgctcg gcctgcgatc tggtcaacgt 2160

ggatctggat gactgtgttt ctgagcaata aatgacttaa accaggtatg gctgccgatg 2220

gttatcttcc agattggctc gaggacaacc tctctgaggg cattcgcgag tggtgggact 2280

tgaaacctgg agccccgaaa cccaaagcca accagcaaaa gcaggacgac ggccggggtc 2340

tggtgcttcc tggctacaag tacctcggac ccttcaacgg actcgacaag ggggagcccg 2400

tcaacgcggc ggatgcagcg gccctcgagc acgacaaggc ctacgaccag cagctcaaag 2460

cgggtgacaa tccgtacctg cggtataacc acgccgacgc cgagtttcag gagcgtctgc 2520

aagaagatac gtcttttggg ggcaacctcg ggcgagcagt cttccaggcc aagaagaggg 2580

ttctcgaacc ttttggtctg gttgaggaag gtgctaagac ggctcctgga aagaaacgtc 2640

cggtagagca gtcgccacaa gagccagact cctcctcggg cattggcaag acaggccagc 2700

agcccgctaa aaagagactc aattttggtc agactggcga ctcagagtca gtccccgacc 2760

cacaacctct cggagaacct ccagcaaccc ccgctgctgt gggacctact acaatggctt 2820

caggcggtgg cgcaccaatg gcagacaata acgaaggcgc cgacggagtg ggtaatgcct 2880

caggaaattg gcattgcgat tccacatggc tgggcgacag agtcatcacc accagcaccc 2940

gaacatgggc cttgcccacc tataacaacc acctctacaa gcaaatctcc agtgcttcaa 3000

cgggggccag caacgacaac cactacttcg gctacagcac cccctggggg tattttgatt 3060
tcaacagatt ccactgccat ttctcaccac gtgactggca gcgactcatc aacaacaatt 3120
ggggattccg gcccaagaga ctcaacttca agctcttcaa catccaagtc aaggaggtca 3180
cgacgaatga tggcgtcacg accatcgcta ataaccttac cagcacggtt caagtcttct 3240
cggactcgga gtaccagttg ccgtacgtcc tcggctctgc gcaccagggc tgcctccctc 3300
cgttcccggc ggacgtgttc atgattccgc agtacggcta cctaacgctc aacaatggca 3360
gccaggcagt gggacggtca tccttttact gcctggaata tttcccatcg cagatgctga 3420
gaacgggcaa taactttacc ttcagctaca ccttcgagga cgtgcctttc cacagcagct 3480
acgcgcacag ccagagcctg gaccggctga tgaatcctct catcgaccag tacctgtatt 3540
acctgaacag aactcagaat cagtccggaa gtgcccaaaa caaggacttg ctgtttagcc 3600
gggggtctcc agctggcatg tctgttcagc ccaaaaactg gctacctgga ccctgttacc 3660
ggcagcagcg cgtttctaaa acaaaaacag acaacaacaa cagcaacttt acctggactg 3720
gtgcttcaaa atataacctt aatgggcgtg aatctataat caaccctggc actgctatgg 3780
cctcacacaa agacgacaaa gacaagttct ttcccatgag cggtgtcatg atttttggaa 3840
aggagagcgc cggagcttca aacactgcat tggacaatgt catgatcaca gacgaagagg 3900
aaatcaaagc cactaacccc gtggccaccg aaagatttgg gactgtggca gtcaatctcc 3960
agagcagcag cacagaccct gcgaccggag atgtgcatgt tatgggagcc ttacctggaa 4020
tggtgtggca agacagagac gtatacctgc agggtcctat ttgggccaaa attcctcaca 4080
cggatggaca ctttcacccg tctcctctca tgggcggctt tggacttaag cacccgcctc 4140
ctcagatcct catcaaaaac acgcctgttc ctgcgaatcc tccggcagag ttttcggcta 4200
caaagtttgc ttcattcatc acccagtatt ccacaggaca agtgagcgtg gagattgaat 4260
gggagctgca gaaagaaaac agcaaacgct ggaatcccga agtgcagtat acatctaact 4320
atgcaaaatc tgccaacgtt gatttcactg tggacaacaa tggactttat actgagcctc 4380
gccccattgg cacccgttac ctcacccgtc ccctgtaatt gtgtgttaat caataaaccg 4440
gttaattcgt gtcagttgaa ctttggtctc atgtcgttat tatcttatct ggtcaccata 4500
gcaaccggtt acacattaac tgcttagttg cgcttcgcga ataccctag tgatggagtt 4560
gcccactccc tctatgcgcg ctcgctcgct cggtggggcc ggcagagcag agctctgccg 4620
tctgcggacc tttggtccgc aggccccacc gagcgagcga gcgcgcatag agggagtggg 4680
caa 4683

<210> 3
<211> 4718
<212> DNA
<213> adeno-associated virus serotype 1

<400> 3
ttgcccactc cctctctgcg cgctcgctcg ctcggtgggg cctgcggacc aaaggtccgc 60

agacggcaga gctctgctct gccggcccca ccgagcgagc gagcgcgcag agagggagtg 120
ggcaactcca tcactagggg taatcgcgaa gcgcctccca cgctgccgcg tcagcgctga 180
cgtaaattac gtcatagggg agtggtcctg tattagctgt cacgtgagtg cttttgcgac 240
attttgcgac accacgtggc catttagggt atatatggcc gagtgagcga gcaggatctc 300
cattttgacc gcgaaatttg aacgagcagc agccatgccg ggcttctacg agatcgtgat 360
caaggtgccg agcgacctgg acgagcacct gccgggcatt tctgactcgt ttgtgagctg 420
ggtggccgag aaggaatggg agctgccccc ggattctgac atggatctga atctgattga 480
gcaggcaccc ctgaccgtgg ccgagaagct gcagcgcgac ttcctggtcc aatggcgccg 540
cgtgagtaag gccccggagg ccctcttctt tgttcagttc gagaagggcg agtcctactt 600
ccacctccat attctggtgg agaccacggg ggtcaaatcc atggtgctgg gccgcttcct 660
gagtcagatt agggacaagc tggtgcagac catctaccgc gggatcgagc cgaccctgcc 720
caactggttc gcggtgacca agacgcgtaa tggcgccgga ggggggaaca aggtggtgga 780
cgagtgctac atccccaact acctcctgcc caagactcag cccgagctgc agtgggcgtg 840
gactaacatg gaggagtata taagcgcctg tttgaacctg gccgagcgca aacggctcgt 900
ggcgcagcac ctgacccacg tcagccagac ccaggagcag aacaaggaga atctgaaccc 960
caattctgac gcgcctgtca tccggtcaaa aacctccgcg cgctacatgg agctggtcgg 1020
gtggctggtg gaccggggca tcacctccga gaagcagtgg atccaggagg accaggcctc 1080
gtacatctcc ttcaacgccg cttccaactc gcggtcccag atcaaggccg ctctggacaa 1140
tgccggcaag atcatggcgc tgaccaaatc cgcgcccgac tacctggtag gccccgctcc 1200
gcccgcggac attaaaacca accgcatcta ccgcatcctg gagctgaacg gctacgaacc 1260
tgcctacgcc ggctccgtct ttctcggctg ggcccagaaa aggttcggga agcgcaacac 1320
catctggctg tttgggccgg ccaccacggg caagaccaac atcgcggaag ccatcgccca 1380
cgccgtgccc ttctacggct gcgtcaactg gaccaatgag aactttccct tcaatgattg 1440
cgtcgacaag atggtgatct ggtgggagga gggcaagatg acggccaagg tcgtggagtc 1500
cgccaaggcc attctcggcg gcagcaaggt gcgcgtggac caaaagtgca agtcgtccgc 1560
ccagatcgac cccacccccg tgatcgtcac ctccaacacc aacatgtgcg ccgtgattga 1620
cgggaacagc accaccttcg agcaccagca gccgttgcag gaccggatgt tcaaatttga 1680
actcacccgc cgtctggagc atgactttgg caaggtgaca aagcaggaag tcaaagagtt 1740
cttccgctgg gcgcaggatc acgtgaccga ggtggcgcat gagttctacg tcagaaaggg 1800
tggagccaac aaaagacccg cccccgatga cgcggataaa agcgagccca agcgggcctg 1860
cccctcagtc gcggatccat cgacgtcaga cgcggaagga gctccggtgg actttgccga 1920
caggtaccaa aacaaatgtt ctcgtcacgc gggcatgctt cagatgctgt ttccctgcaa 1980
gacatgcgag agaatgaatc agaatttcaa catttgcttc acgcacggga cgagagactg 2040

ttcagagtgc ttccccggcg tgtcagaatc tcaaccggtc gtcagaaaga ggacgtatcg 2100

gaaactctgt gccattcatc atctgctggg gcgggctccc gagattgctt gctcggcctg 2160

cgatctggtc aacgtggacc tggatgactg tgtttctgag caataaatga cttaaaccag 2220

gtatggctgc cgatggttat cttccagatt ggctcgagga caacctctct gagggcattc 2280

gcgagtggtg ggacttgaaa cctggagccc cgaagcccaa agccaaccag caaaagcagg 2340

acgacggccg gggtctggtg cttcctggct acaagtacct cggacccttc aacggactcg 2400

acaaggggga gcccgtcaac gcggcggacg cagcggccct cgagcacgac aaggcctacg 2460

accagcagct caaagcgggt gacaatccgt acctgcggta taaccacgcc gacgccgagt 2520

ttcaggagcg tctgcaagaa gatacgtctt ttgggggcaa cctcgggcga gcagtcttcc 2580

aggccaagaa gcgggttctc gaacctctcg gtctggttga ggaaggcgct aagacggctc 2640

ctggaaagaa acgtccggta gagcagtcgc cacaagagcc agactcctcc tcgggcatcg 2700

gcaagacagg ccagcagccc gctaaaaaga gactcaattt tggtcagact ggcgactcag 2760

agtcagtccc cgatccacaa cctctcggag aacctccagc aacccccgct gctgtgggac 2820

ctactacaat ggcttcaggc ggtggcgcac caatggcaga caataacgaa ggcgccgacg 2880

gagtgggtaa tgcctcagga aattggcatt gcgattccac atggctgggc gacagagtca 2940

tcaccaccag cacccgcacc tgggccttgc ccacctacaa taaccacctc tacaagcaaa 3000

tctccagtgc ttcaacgggg gccagcaacg acaaccacta cttcggctac agcacccccc 3060

gggggtattt tgatttcaac agattccact gccacttttc accacgtgac tggcagcgac 3120

tcatcaacaa caattgggga ttccggccca agagactcaa cttcaaactc ttcaacatcc 3180

aagtcaagga ggtcacgacg aatgatggcg tcacaaccat cgctaataac cttaccagca 3240

cggttcaagt cttctcggac tcggagtacc agcttccgta cgtcctcggc tctgcgcacc 3300

agggctgcct ccctccgttc ccggcggacg tgttcatgat tccgcaatac ggctacctga 3360

cgctcaacaa tggcagccaa gccgtgggac gttcatcctt ttactgcctg gaatatttcc 3420

cttctcagat gctgagaacg ggcaacaact ttaccttcag ctacaccttt gaggaagtgc 3480

ctttccacag cagctacgcg cacagccaga gcctggaccg gctgatgaat cctctcatcg 3540

accaatacct gtattacctg aacagaactc aaaatcagtc cggaagtgcc caaaacaagg 3600

acttgctgtt tagccgtggg tctccagctg gcatgtctgt tcagcccaaa aactggctac 3660

ctggaccctg ttatcggcag cagcgcgttt ctaaaacaaa aacagacaac aacaacagca 3720

attttacctg gactggtgct tcaaaatata acctcaatgg gcgtgaatcc atcatcaacc 3780

ctggcactgc tatggcctca cacaaagacg acgaagacaa gttctttccc atgagcggtg 3840

tcatgatttt tggaaaagag agcgccggag cttcaaacac tgcattggac aatgtcatga 3900

ttacagacga agaggaaatt aaagccacta accctgtggc caccgaaaga tttgggaccg 3960

tggcagtcaa tttccagagc agcagcacag accctgcgac cggagatgtg catgctatgg 4020

gagcattacc tggcatggtg tggcaagata gagacgtgta cctgcagggt cccatttggg 4080

ccaaaattcc tcacacagat ggacactttc acccgtctcc tcttatgggc ggctttggac 4140

tcaagaaccc gcctcctcag atcctcatca aaaacacgcc tgttcctgcg aatcctccgg 4200

cggagttttc agctacaaag tttgcttcat tcatcaccca atactccaca ggacaagtga 4260

gtgtggaaat tgaatgggag ctgcagaaag aaaacagcaa gcgctggaat cccgaagtgc 4320

agtacacatc caattatgca aaatctgcca acgttgattt tactgtggac aacaatggac 4380

tttatactga gcctcgcccc attggcaccc gttaccttac ccgtcccctg taattacgtg 4440

ttaatcaata aaccggttga ttcgtttcag ttgaactttg gtctcctgtc cttcttatct 4500

tatcggttac catggttata gcttacacat taactgcttg gttgcgcttc gcgataaaag 4560

acttacgtca tcgggttacc cctagtgatg gagttgccca ctccctctct gcgcgctcgc 4620

tcgctcggtg gggcctgcgg accaaaggtc cgcagacggc agagctctgc tctgccggcc 4680

ccaccgagcg agcgagcgcg cagagaggga gtgggcaa 4718

<210> 4
<211> 4726
<212> DNA
<213> adeno-associated virus serotype 3

<400> 4
ttggccactc cctctatgcg cactcgctcg ctcggtgggg cctggcgacc aaaggtcgcc 60

agacggacgt gctttgcacg tccggcccca ccgagcgagc gagtgcgcat agagggagtg 120

gccaactcca tcactagagg tatggcagtg acgtaacgcg aagcgcgcga agcgagacca 180

cgcctaccag ctgcgtcagc agtcaggtga ccctttgcg acagtttgcg acaccacgtg 240

gccgctgagg gtatatattc tcgagtgagc gaaccaggag ctccattttg accgcgaaat 300

ttgaacgagc agcagccatg ccggggttct acgagattgt cctgaaggtc ccgagtgacc 360

tggacgagcg cctgccgggc atttctaact cgtttgttaa ctgggtggcc gagaaggaat 420

gggacgtgcc gccggattct gacatggatc cgaatctgat tgagcaggca cccctgaccg 480

tggccgaaaa gcttcagcgc gagttcctgg tggagtggcg ccgcgtgagt aaggccccgg 540

aggccctctt ttttgtccag ttcgaaaagg gggagaccta cttccacctg cacgtgctga 600

ttgagaccat cggggtcaaa tccatggtgg tcggccgcta cgtgagccag attaaagaga 660

agctggtgac ccgcatctac cgcggggtcg agccgcagct tccgaactgg ttcgcggtga 720

ccaaaacgcg aaatggcgcc gggggcggga acaaggtggt ggacgactgc tacatcccca 780

actacctgct ccccaagacc cagcccgagc tccagtgggc gtggactaac atggaccagt 840

atttaagcgc ctgtttgaat ctcgcggagc gtaaacggct ggtggcgcag catctgacgc 900

acgtgtcgca gacgcaggag cagaacaaag agaatcagaa ccccaattct gacgcgccgg 960

tcatcaggtc aaaaacctca gccaggtaca tggagctggt cgggtggctg gtggaccgcg 1020

ggatcacgtc agaaaagcaa tggattcagg aggaccaggc ctcgtacatc tccttcaacg 1080

ccgcctccaa ctcgcggtcc cagatcaagg ccgcgctgga caatgcctcc aagatcatga 1140

gcctgacaaa gacggctccg gactacctgg tgggcagcaa cccgccggag gacattacca 1200
aaaatcggat ctaccaaatc ctggagctga acgggtacga tccgcagtac gcggcctccg 1260
tcttcctggg ctgggcgcaa aagaagttcg ggaagaggaa caccatctgg ctctttgggc 1320
cggccacgac gggtaaaacc aacatcgcgg aagccatcgc ccacgccgtg cccttctacg 1380
gctgcgtaaa ctggaccaat gagaactttc ccttcaacga ttgcgtcgac aagatggtga 1440
tctggtggga ggagggcaag atgacggcca aggtcgtgga gagcgccaag gccattctgg 1500
gcggaagcaa ggtgcgcgtg gaccaaaagt gcaagtcatc ggcccagatc gaacccactc 1560
ccgtgatcgt cacctccaac accaacatgt gcgccgtgat tgacgggaac agcaccacct 1620
tcgagcatca gcagccgctg caggaccgga tgtttgaatt tgaacttacc cgccgtttgg 1680
accatgactt tgggaaggtc accaaacagg aagtaaagga ctttttccgg tgggcttccg 1740
atcacgtgac tgacgtggct catgagttct acgtcagaaa gggtggagct aagaaacgcc 1800
ccgcctccaa tgacgcggat gtaagcgagc caaaacggga gtgcacgtca cttgcgcagc 1860
cgacaacgtc agacgcggaa gcaccggcgg actacgcgga caggtaccaa aacaaatgtt 1920
ctcgtcacgt gggcatgaat ctgatgcttt ttccctgtaa aacatgcgag agaatgaatc 1980
aaatttccaa tgtctgtttt acgcatggtc aaagagactg tggggaatgc ttccctggaa 2040
tgtcagaatc tcaacccgtt tctgtcgtca aaaagaagac ttatcagaaa ctgtgtccaa 2100
ttcatcatat cctgggaagg gcacccgaga ttgcctgttc ggcctgcgat ttggccaatg 2160
tggacttgga tgactgtgtt tctgagcaat aaatgactta aaccaggtat ggctgctgac 2220
ggttatcttc cagattggct cgaggacaac ctttctgaag gcattcgtga gtggtgggct 2280
ctgaaacctg gagtccctca acccaaagcg aaccaacaac accaggacaa ccgtcggggt 2340
cttgtgcttc cgggttacaa atacctcgga cccggtaacg gactcgacaa aggagagccg 2400
gtcaacgagg cggacgcggc agccctcgaa cacgacaaag cttacgacca gcagctcaag 2460
gccggtgaca acccgtacct caagtacaac cacgccgacg ccgagtttca ggagcgtctt 2520
caagaagata cgtcttttgg gggcaacctt ggcagagcag tcttccaggc caaaaagagg 2580
atccttgagc ctcttggtct ggttgaggaa gcagctaaaa cggctcctgg aaagaagggg 2640
gctgtagatc agtctcctca ggaaccggac tcatcatctg gtgttggcaa atcgggcaaa 2700
cagcctgcca gaaaaagact aaatttcggt cagactggag actcagagtc agtcccagac 2760
cctcaacctc tcggagaacc accagcagcc cccacaagtt tgggatctaa tacaatggct 2820
tcaggcggtg gcgcaccaat ggcagacaat aacgagggtg ccgatggagt gggtaattcc 2880
tcaggaaatt ggcattgcga ttcccaatgg ctgggcgaca gagtcatcac caccagcacc 2940
agaacctggg ccctgcccac ttacaacaac catctctaca agcaaatctc cagccaatca 3000
ggagcttcaa acgacaacca ctactttggc tacagcaccc cttgggggta ttttgacttt 3060
aacagattcc actgccactt ctcaccacgt gactggcagc gactcattaa caacaactgg 3120

ggattccggc ccaagaaact cagcttcaag ctcttcaaca tccaagttag aggggtcacg 3180

cagaacgatg gcacgacgac tattgccaat aaccttacca gcacggttca agtgtttacg 3240

gactcggagt atcagctccc gtacgtgctc gggtcggcgc accaaggctg tctcccgccg 3300

tttccagcgg acgtcttcat ggtccctcag tatggatacc tcaccctgaa caacggaagt 3360

caagcggtgg gacgctcatc cttttactgc ctggagtact tcccttcgca gatgctaagg 3420

actggaaata acttccaatt cagctatacc ttcgaggatg taccttttca cagcagctac 3480

gctcacagcc agagtttgga tcgcttgatg aatcctctta ttgatcagta tctgtactac 3540

ctgaacagaa cgcaaggaac aacctctgga acaaccaacc aatcacggct gctttttagc 3600

caggctgggc ctcagtctat gtctttgcag gccagaaatt ggctacctgg gccctgctac 3660

cggcaacaga gactttcaaa gactgctaac gacaacaaca acagtaactt tccttggaca 3720

gcggccagca aatatcatct caatggccgc gactcgctgg tgaatccagg accagctatg 3780

gccagtcaca aggacgatga agaaaaattt ttccctatgc acggcaatct aatatttggc 3840

aaagaaggga caacggcaag taacgcagaa ttagataatg taatgattac ggatgaagaa 3900

gagattcgta ccaccaatcc tgtggcaaca gagcagtatg gaactgtggc aaataacttg 3960

cagagctcaa atacagctcc cacgactgga actgtcaatc atcaggggc cttacctggc 4020

atggtgtggc aagatcgtga cgtgtacctt caaggaccta tctgggcaaa gattcctcac 4080

acggatggac actttcatcc ttctcctctg atgggaggct ttggactgaa acatccgcct 4140

cctcaaatca tgatcaaaaa tactccggta ccggcaaatc ctccgacgac tttcagcccg 4200

gccaagtttg cttcatttat cactcagtac tccactggac aggtcagcgt ggaaattgag 4260

tgggagctac agaaagaaaa cagcaaacgt tggaatccag agattcagta cacttccaac 4320

tacaacaagt ctgttaatgt ggactttact gtagacacta atggtgttta tagtgaacct 4380

cgccctattg gaacccggta tctcacacga aacttgtgaa tcctggttaa tcaataaacc 4440

gtttaattcg tttcagttga actttggctc ttgtgcactt ctttatcttt atcttgtttc 4500

catggctact gcgtagataa gcagcggcct gcggcgcttg cgcttcgcgg tttacaactg 4560

ctggttaata tttaactctc gccatacctc tagtgatgga gttggccact ccctctatgc 4620

gcactcgctc gctcggtggg gcctggcgac caaaggtcgc cagacggacg tgctttgcac 4680

gtccggcccc accgagcgag cgagtgcgca tagagggagt ggccaa 4726

<210> 5
<211> 4767
<212> DNA
<213> adeno-associated virus serotype 4

<400> 5
ttggccactc cctctatgcg cgctcgctca ctcactcggc cctggagacc aaaggtctcc 60

agactgccgg cctctggccg gcagggccga gtgagtgagc gagcgcgcat agagggagtg 120

gccaactcca tcatctaggt ttgcccactg acgtcaatgt gacgtcctag ggttagggag 180

gtccctgtat tagcagtcac gtgagtgtcg tatttcgcgg agcgtagcgg agcgcatacc 240

aagctgccac gtcacagcca cgtggtccgt ttgcgacagt ttgcgacacc atgtggtcag 300

gagggtatat aaccgcgagt gagccagcga ggagctccat tttgcccgcg aattttgaac 360

gagcagcagc catgccgggg ttctacgaga tcgtgctgaa ggtgcccagc gacctggacg 420

agcacctgcc cggcatttct gactcttttg tgagctgggt ggccgagaag gaatgggagc 480

tgccgccgga ttctgacatg gacttgaatc tgattgagca ggcacccctg accgtggccg 540

aaaagctgca acgcgagttc ctggtcgagt ggcgccgcgt gagtaaggcc ccggaggccc 600

tcttctttgt ccagttcgag aaggggggaca gctacttcca cctgcacatc ctggtggaga 660

ccgtgggcgt caaatccatg gtggtgggcc gctacgtgag ccagattaaa gagaagctgg 720

tgacccgcat ctaccgcggg gtcgagccgc agcttccgaa ctggttcgcg gtgaccaaga 780

cgcgtaatgg cgccggaggc gggaacaagg tggtggacga ctgctacatc cccaactacc 840

tgctccccaa gacccagccc gagctccagt gggcgtggac taacatggac cagtatataa 900

gcgcctgttt gaatctcgcg gagcgtaaac ggctggtggc gcagcatctg acgcacgtgt 960

cgcagacgca ggagcagaac aaggaaaacc agaaccccaa ttctgacgcg ccggtcatca 1020

ggtcaaaaac ctccgccagg tacatggagc tggtcgggtg gctggtggac cgcgggatca 1080

cgtcagaaaa gcaatggatc caggaggacc aggcgtccta catctccttc aacgccgcct 1140

ccaactcgcg gtcacaaatc aaggccgcgc tggacaatgc ctccaaaatc atgagcctga 1200

caaagacggc tccggactac ctggtgggcc agaacccgcc ggaggacatt tccagcaacc 1260

gcatctaccg aatcctcgag atgaacgggt acgatccgca gtacgcggcc tccgtcttcc 1320

tgggctgggc gcaaaagaag ttcgggaaga ggaacaccat ctggctcttt gggccggcca 1380

cgacgggtaa aaccaacatc gcggaagcca tcgcccacgc cgtgcccttc tacggctgcg 1440

tgaactggac caatgagaac tttccgttca acgattgcgt cgacaagatg gtgatctggt 1500

gggaggaggg caagatgacg gccaaggtcg tagagagcgc caaggccatc ctgggcggaa 1560

gcaaggtgcg cgtggaccaa aagtgcaagt catcggccca gatcgaccca actcccgtga 1620

tcgtcacctc caacaccaac atgtgcgcgg tcatcgacgg aaactcgacc accttcgagc 1680

accaacaacc actccaggac cggatgttca agttcgagct caccaagcgc ctggagcacg 1740

actttggcaa ggtcaccaag caggaagtca aagacttttt ccggtgggcg tcagatcacg 1800

tgaccgaggt gactcacgag ttttacgtca gaaagggtgg agctagaaag aggcccgccc 1860

ccaatgacgc agatataagt gagcccaagc gggcctgtcc gtcagttgcg cagccatcga 1920

cgtcagacgc ggaagctccg gtggactacg cggacaggta ccaaaacaaa tgttctcgtc 1980

acgtgggtat gaatctgatg ctttttccct gccggcaatg cgagagaatg aatcagaatg 2040

tggacatttg cttcacgcac ggggtcatgg actgtgccga gtgcttcccc gtgtcagaat 2100

ctcaacccgt gtctgtcgtc agaaagcgga cgtatcagaa actgtgtccg attcatcaca 2160

tcatggggag ggcgcccgag gtggcctgct cggcctgcga actggccaat gtggacttgg 2220

atgactgtga catggaacaa taaatgactc aaaccagata tgactgacgg ttaccttcca 2280
gattggctag aggacaacct ctctgaaggc gttcgagagt ggtgggcgct gcaacctgga 2340
gcccctaaac ccaaggcaaa tcaacaacat caggacaacg ctcggggtct tgtgcttccg 2400
ggttacaaat acctcggacc cggcaacgga ctcgacaagg gggaacccgt caacgcagcg 2460
gacgcggcag ccctcgagca cgacaaggcc tacgaccagc agctcaaggc cggtgacaac 2520
ccctacctca agtacaacca cgccgacgcg gagttccagc agcggcttca gggcgacaca 2580
tcgtttgggg gcaacctcgg cagagcagtc ttccaggcca aaaagagggt tcttgaacct 2640
cttggtctgg ttgagcaagc gggtgagacg gctcctggaa agaagagacc gttgattgaa 2700
tccccccagc agcccgactc ctccacgggt atcggcaaaa aaggcaagca gccggctaaa 2760
aagaagctcg ttttcgaaga cgaaactgga gcaggcgacg gacccccctga gggatcaact 2820
tccggagcca tgtctgatga cagtgagatg cgtgcagcag ctggcggagc tgcagtcgag 2880
ggcggacaag gtgccgatgg agtgggtaat gcctcgggtg attggcattg cgattccacc 2940
tggtctgagg gccacgtcac gaccaccagc accagaacct gggtcttgcc cacctacaac 3000
aaccacctct acaagcgact cggagagagc ctgcagtcca acacctacaa cggattctcc 3060
acccccctggg gatactttga cttcaaccgc ttccactgcc acttctcacc acgtgactgg 3120
cagcgactca tcaacaacaa ctggggcatg cgacccaaag ccatgcgggt caaaatcttc 3180
aacatccagg tcaaggaggt cacgacgtcg aacggcgaga caacggtggc taataacctt 3240
accagcacgg ttcagatctt tgcggactcg tcgtacgaac tgccgtacgt gatggatgcg 3300
ggtcaagagg gcagcctgcc tccttttccc aacgacgtct ttatggtgcc ccagtacggc 3360
tactgtggac tggtgaccgg caacacttcg cagcaacaga ctgacagaaa tgccttctac 3420
tgcctggagt actttccttc gcagatgctg cggactggca acaactttga aattacgtac 3480
agttttgaga aggtgccttt ccactcgatg tacgcgcaca gccagagcct ggaccggctg 3540
atgaaccctc tcatcgacca gtacctgtgg ggactgcaat cgaccaccac cggaaccacc 3600
ctgaatgccg ggactgccac caccaacttt accaagctgc ggcctaccaa cttttccaac 3660
tttaaaaaga actggctgcc cgggccttca atcaagcagc agggcttctc aaagactgcc 3720
aatcaaaact acaagatccc tgccaccggg tcagacagtc tcatcaaata cgagacgcac 3780
agcactctgg acggaagatg gagtgccctg acccccggac ctccaatggc cacggctgga 3840
cctgcggaca gcaagttcag caacagccag ctcatctttg cggggcctaa acagaacggc 3900
aacacggcca ccgtacccgg gactctgatc ttcacctctg aggaggagct ggcagccacc 3960
aacgccaccg atacggacat gtggggcaac ctacctggcg gtgaccagag caacagcaac 4020
ctgccgaccg tggacagact gacagccttg ggagccgtgc ctggaatggt ctggcaaaac 4080
agagacattt actaccaggg tcccatttgg gccaagattc ctcataccga tggacacttt 4140
cacccctcac cgctgattgg tgggtttggg ctgaaacacc cgcctcctca aatttttatc 4200

aagaacaccc cggtacctgc gaatcctgca acgaccttca gctctactcc ggtaaactcc 4260

ttcattactc agtacagcac tggccaggtg tcggtgcaga ttgactggga gatccagaag 4320

gagcggtcca aacgctggaa ccccgaggtc cagtttacct ccaactacgg acagcaaaac 4380

tctctgttgt gggctcccga tgcggctggg aaatacactg agcctagggc tatcggtacc 4440

cgctacctca cccaccacct gtaataacct gttaatcaat aaaccggttt attcgtttca 4500

gttgaacttt ggtctccgtg tccttcttat cttatctcgt ttccatggct actgcgtaca 4560

taagcagcgg cctgcggcgc ttgcgcttcg cggtttacaa ctgccggtta atcagtaact 4620

tctggcaaac cagatgatgg agttggccac attagctatg cgcgctcgct cactcactcg 4680

gccctggaga ccaaaggtct ccagactgcc ggcctctggc cggcagggcc gagtgagtga 4740

gcgagcgcgc atagagggag tggccaa 4767

<210> 6
<211> 4404
<212> DNA
<213> adeno-associated virus serotype 5

<400> 6
cgcgacaggg gggagagtgc cacactctca agcaaggggg ttttgtaagc agtgatgtca 60

taatgatgta atgcttattg tcacgcgata gttaatgatt aacagtcatg tgatgtgttt 120

tatccaatag gaagaaagcg cgcgtatgag ttctcgcgag acttccgggg tataaaagac 180

cgagtgaacg agcccgccgc cattctttgc tctggactgc tagaggaccc tcgctgccat 240

ggctaccttc tatgaagtca ttgttcgcgt cccatttgac gtggaggaac atctgcctgg 300

aatttctgac agctttgtgg actgggtaac tggtcaaatt tgggagctgc ctccagagtc 360

agatttaaat ttgactctgg ttgaacagcc tcagttgacg gtggctgata gaattcgccg 420

cgtgttcctg tacgagtgga acaaattttc caagcaggag tccaaattct ttgtgcagtt 480

tgaaaaggga tctgaatatt ttcatctgca cacgcttgtg gagacctccg gcatctcttc 540

catggtcctc ggccgctacg tgagtcagat tcgcgcccag ctggtgaaag tggtcttcca 600

gggaattgaa ccccagatca acgactgggt cgccatcacc aaggtaaaga agggcggagc 660

caataaggtg gtggattctg ggtatattcc cgcctacctg ctgccgaagg tccaaccgga 720

gcttcagtgg gcgtggacaa acctggacga gtataaattg gccgccctga atctggagga 780

gcgcaaacgg ctcgtcgcgc agtttctggc agaatcctcg cagcgctcgc aggaggcggc 840

ttcgcagcgt gagttctcgg ctgacccggt catcaaaagc aagacttccc agaaatacat 900

ggcgctcgtc aactggctcg tggagcacgg catcacttcc gagaagcagt ggatccagga 960

aaatcaggag agctacctct ccttcaactc caccggcaac tctcggagcc agatcaaggc 1020

cgcgctcgac aacgcgacca aaattatgag tctgacaaaa agcgcggtgg actacctcgt 1080

ggggagctcc gttcccgagg acatttcaaa aaacagaatc tggcaaattt ttgagatgaa 1140

tggctacgac ccggcctacg cgggatccat cctctacggc tggtgtcagc gctccttcaa 1200

caagaggaac accgtctggc tctacggacc cgccacgacc ggcaagacca acatcgcgga 1260

ggccatcgcc cacactgtgc ccttttacgg ctgcgtgaac tggaccaatg aaaactttcc 1320

ctttaatgac tgtgtggaca aaatgctcat ttggtgggag gagggaaaga tgaccaacaa 1380

ggtggttgaa tccgccaagg ccatcctggg gggctcaaag gtgcgggtcg atcagaaatg 1440

taaatcctct gttcaaattg attctacccc tgtcattgta acttccaata caaacatgtg 1500

tgtggtggtg gatgggaatt ccacgacctt tgaacaccag cagccgctgg aggaccgcat 1560

gttcaaattt gaactgacta agcggctccc gccagatttt ggcaagatta ctaagcagga 1620

agtcaaggac ttttttgctt gggcaaaggt caatcaggtg ccggtgactc acgagtttaa 1680

agttcccagg gaattggcgg gaactaaagg ggcggagaaa tctctaaaac gcccactggg 1740

tgacgtcacc aatactagct ataaaagtct ggagaagcgg gccaggctct catttgttcc 1800

cgagacgcct cgcagttcag acgtgactgt tgatcccgct cctctgcgac cgctcaattg 1860

gaattcaagg tatgattgca aatgtgacta tcatgctcaa tttgacaaca tttctaacaa 1920

atgtgatgaa tgtgaatatt tgaatcgggg caaaaatgga tgtatctgtc acaatgtaac 1980

tcactgtcaa atttgtcatg ggattccccc ctgggaaaag gaaaacttgt cagattttgg 2040

ggattttgac gatgccaata aagaacagta aataaagcga gtagtcatgt cttttgttga 2100

tcaccctcca gattggttgg aagaagttgg tgaaggtctt cgcgagtttt tgggccttga 2160

agcgggccca ccgaaaccaa aacccaatca gcagcatcaa gatcaagccc gtggtcttgt 2220

gctgcctggt tataactatc tcggacccgg aaacggtctc gatcgaggag agcctgtcaa 2280

cagggcagac gaggtcgcgc gagagcacga catctcgtac aacgagcagc ttgaggcggg 2340

agacaacccc tacctcaagt acaaccacgc ggacgccgag tttcaggaga agctcgccga 2400

cgacacatcc ttcgggggaa acctcggaaa ggcagtcttt caggccaaga aaagggttct 2460

cgaacctttt ggcctggttg aagagggtgc taagacggcc cctaccggaa agcggataga 2520

cgaccacttt ccaaaaagaa agaaggctcg gaccgaagag gactccaagc cttccacctc 2580

gtcagacgcc gaagctggac ccagcggatc ccagcagctg caaatcccag cccaaccagc 2640

ctcaagtttg ggagctgata caatgtctgc gggaggtggc ggcccattgg gcgacaataa 2700

ccaaggtgcc gatggagtgg gcaatgcctc gggagattgg cattgcgatt ccacgtggat 2760

gggggacaga gtcgtcacca agtccacccg aacctgggtg ctgcccagct acaacaacca 2820

ccagtaccga gagatcaaaa gcggctccgt cgacggaagc aacgccaacg cctactttgg 2880

atacagcacc ccctgggggt actttgactt taaccgcttc cacagccact ggagcccccg 2940

agactggcaa agactcatca acaactactg gggcttcaga ccccggtccc tcagagtcaa 3000

aatcttcaac attcaagtca aagaggtcac ggtgcaggac tccaccacca ccatcgccaa 3060

caacctcacc tccaccgtcc aagtgtttac ggacgacgac taccagctgc cctacgtcgt 3120

cggcaacggg accgagggat gcctgccggc cttccctccg caggtcttta cgctgccgca 3180

gtacggttac gcgacgctga accgcgacaa cacagaaaat cccaccgaga ggagcagctt 3240

cttctgccta gagtactttc ccagcaagat gctgagaacg ggcaacaact ttgagtttac 3300

ctacaacttt gaggaggtgc ccttccactc cagcttcgct cccagtcaga acctcttcaa 3360

gctggccaac ccgctggtgg accagtactt gtaccgcttc gtgagcacaa ataacactgg 3420

cggagtccag ttcaacaaga acctggccgg gagatacgcc aacacctaca aaaactggtt 3480

cccggggccc atgggccgaa cccagggctg gaacctgggc tccggggtca accgcgccag 3540

tgtcagcgcc ttcgccacga ccaataggat ggagctcgag ggcgcgagtt accaggtgcc 3600

cccgcagccg aacggcatga ccaacaacct ccagggcagc aacacctatg ccctggagaa 3660

cactatgatc ttcaacagcc agccggcgaa cccgggcacc accgccacgt acctcgaggg 3720

caacatgctc atcaccagcg agagcgagac gcagccggtg aaccgcgtgg cgtacaacgt 3780

cggcgggcag atggccacca acaaccagag ctccaccact gcccccgcga ccggcacgta 3840

caacctccag gaaatcgtgc ccggcagcgt gtggatggag agggacgtgt acctccaagg 3900

acccatctgg gccaagatcc cagagacggg ggcgcacttt caccctctc cggccatggg 3960

cggattcgga ctcaaacacc caccgcccat gatgctcatc aagaacacgc ctgtgcccgg 4020

aaatatcacc agcttctcgg acgtgcccgt cagcagcttc atcacccagt acagcaccgg 4080

gcaggtcacc gtggagatgg agtgggagct caagaaggaa aactccaaga ggtggaaccc 4140

agagatccag tacacaaaca actacaacga cccccagttt gtggactttg ccccggacag 4200

caccggggaa tacagaacca ccagacctat cggaacccga taccttaccc gaccccttta 4260

acccattcat gtcgcatacc ctcaataaac cgtgtattcg tgtcagtaaa atactgcctc 4320

ttgtggtcat tcaatgaata acagcttaca acatctacaa aacctccttg cttgagagtg 4380

tggcactctc ccccctgtcg cgcg 4404