



US012383587B2

(12) **United States Patent**
Choi et al.

(10) **Patent No.:** US 12,383,587 B2
(b5) **Date of Patent:** Aug. 12, 2025

(54) **AAV TRIPLE-PLASMID SYSTEM**(71) Applicant: **TAKEDA PHARMACEUTICAL COMPANY LIMITED**, Osaka (JP)(72) Inventors: **Vivian Choi**, Lexington, MA (US); **Xing Li**, Jamaica Plain, MA (US)(73) Assignee: **TAKEDA PHARMACEUTICAL COMPANY LIMITED**, Osaka (JP)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 920 days.

(21) Appl. No.: **17/288,141**(22) PCT Filed: **Oct. 24, 2019**(86) PCT No.: **PCT/US2019/057916**

§ 371 (c)(1),

(2) Date: **Apr. 23, 2021**(87) PCT Pub. No.: **WO2020/086881**PCT Pub. Date: **Apr. 30, 2020**(65) **Prior Publication Data**

US 2021/0275614 A1 Sep. 9, 2021

Related U.S. Application Data

(60) Provisional application No. 62/750,603, filed on Oct. 25, 2018.

(51) **Int. Cl.***A61K 35/761* (2015.01)*A61K 38/17* (2006.01)*C12N 15/11* (2006.01)(52) **U.S. Cl.**CPC *A61K 35/761* (2013.01); *A61K 38/17* (2013.01); *C12N 15/11* (2013.01)(58) **Field of Classification Search**CPC A61K 35/761; A61K 38/17
See application file for complete search history.(56) **References Cited**

U.S. PATENT DOCUMENTS

9,840,719 B2	12/2017	High et al.
2005/0112765 A1	5/2005	Li
2016/0032319 A1*	2/2016	Wright A61K 48/0091 435/235.1
2016/0108373 A1	4/2016	Bennett et al.
2016/0222356 A1	8/2016	Zhao et al.
2016/0273058 A1	9/2016	Akashika et al.
2017/0021038 A1*	1/2017	Pan A61P 27/02

(Continued)

FOREIGN PATENT DOCUMENTS

CN	104470545	3/2015
CN	105612253	5/2016

(Continued)

OTHER PUBLICATIONS

Robert et al. Manufacturing of recombinant adeno-associated viruses using mammalian expression platforms. Biotechnology J. 12:1-16. (Year: 2017).*

(Continued)

Primary Examiner — Arthur S Leonard

Assistant Examiner — Keenan A Bates

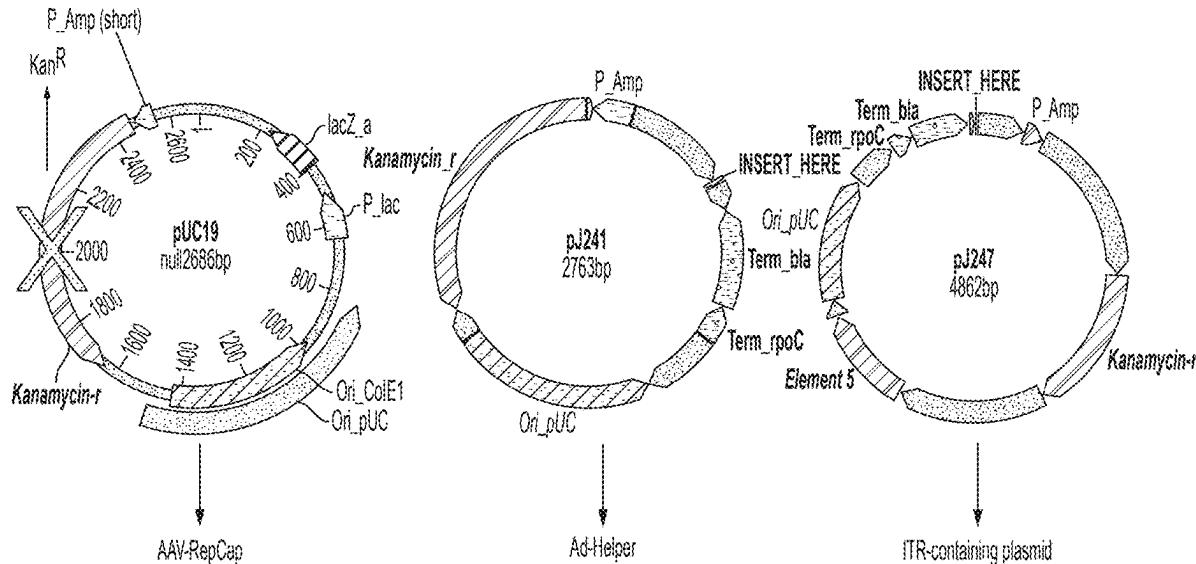
(74) Attorney, Agent, or Firm — Troutman Pepper Locke LLP

(57) **ABSTRACT**

A triple-plasmid system for producing recombinant adeno-associated viruses is disclosed.

12 Claims, 17 Drawing Sheets

Specification includes a Sequence Listing.



(56)

References Cited**U.S. PATENT DOCUMENTS**

2018/0057842 A1 3/2018 Stanley et al.
2019/0255152 A1 8/2019 Kim et al.

FOREIGN PATENT DOCUMENTS

CN	105765066	7/2016
JP	4473346 B2	6/2010
JP	2016517278 A	6/2016
JP	2016-525356	8/2016
JP	2018515102 A	6/2018
WO	2012158757 A1	11/2012
WO	2013164793 A2	11/2013
WO	2014144486 A2	9/2014
WO	2015031686 A1	3/2015
WO	2016139321 A1	9/2016
WO	2016196507 A1	12/2016
WO	2018080277	5/2017
WO	2019195727 A1	10/2019

OTHER PUBLICATIONS

Addgene (pLA230 Sequence Analyzer; sequence from Shinkai A et al. <https://www.addgene.org/browse/sequence/7054/> (Year: 2024).*
Shinkai A et al. In vivo mutagenesis by *Escherichia coli* DNA polymerase I. Ile(709) in motif A functions in base selection. *J Biol Chem.* 276(50):46759-64 (Year: 2001).*
Trapani et al. Methods in Molecular Biology 1715: Retinal Gene Therapy. Chapter 11: Dual AAV Vectors for Stargardt Disease: 153-175. (Year: 2017).*
Epitope-tagged transcription factor ChIP-seq (pp. 1-4. 2017). https://www.encodeproject.org/documents/35a9f776-dd6a-44e3-8795-50ead83f34f7/@/download/attachment/Guidelines_for_Use_of_Epitope_Tags_in_ChIP-seq_Jan_2017.pdf (Year: 2017).*

Ranjan et al. FLAG tag module for PCR based gene targeting. *Journal of Yeast and Fungal Research* 1: 165-169. (Year: 2010).* European Search Report issued Jul. 11, 2022 in connection with EP Application No. 19876857.4.

Xiao et al., "Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus", *Journal of Virology*, the American Society for Microbiology, US, vol. 72, No. 3, Mar. 1, 1998, pp. 2224-2232.

English Translation and Japan Office Action for Japanese Patent Application No. 2021-516750 (dated Sep. 22, 2023).

International Search Report mailed Jan. 27, 2020 in connection with PCT/US19/57916.

Written Opinion mailed Jan. 27, 2020 in connection with PCT/US19/57916.

Wang, et al., "A reliable and feasible qPCR strategy for titrating AAV vectors." *Medical Science Monitor—Basic Research*, Jul. 5, 2013 (Jul. 5, 2013), vol. 19, pp. 187-193.

English Translation and Office Action for Chinese Patent Application No. 201980069110.4 (dated Aug. 17, 2023).

Wright, JF, "Manufacturing and Characterizing AAV-based Vector For Use in Clinical Studies," *Gene Therapy* 15:840-848 (2008).

Bennicelli et al., "Reversal of Blindness in Animal Models of Leber Congenital Amaurosis Using Optimized AAV2-mediated Gene Transfer," *Molecular Therapy* 16(3):458-465 (2008).

Hauck et al., "Undetectable Transcription of cap in a Clinical AAV Vector: Implications for Performed Capsid in Immune Responses," *Molecular Therapy* 17(1):144-152 (2009).

Smith et al., "Packaging of Host Cell and Plasmid DNA into Recombinant Adeno-Associated Virus Particles Produced by Triple Transfection," *Molecular Therapy* 7(5):S348 (2003).

Ozawa, K., "Gene Therapy Using AAV Vectors," *Drug Delivery System* 22-6, p. 643-650 (2007) (English abstract only).

Chadeuf et al., "Evidence for Encapsulation of Prokaryotic Sequences During Recombinant Adeno-Associated Virus Production and Their in Vivo Persistence After Vector Delivery," *Molecular Therapy* 12(4):744-753 (2005).

* cited by examiner

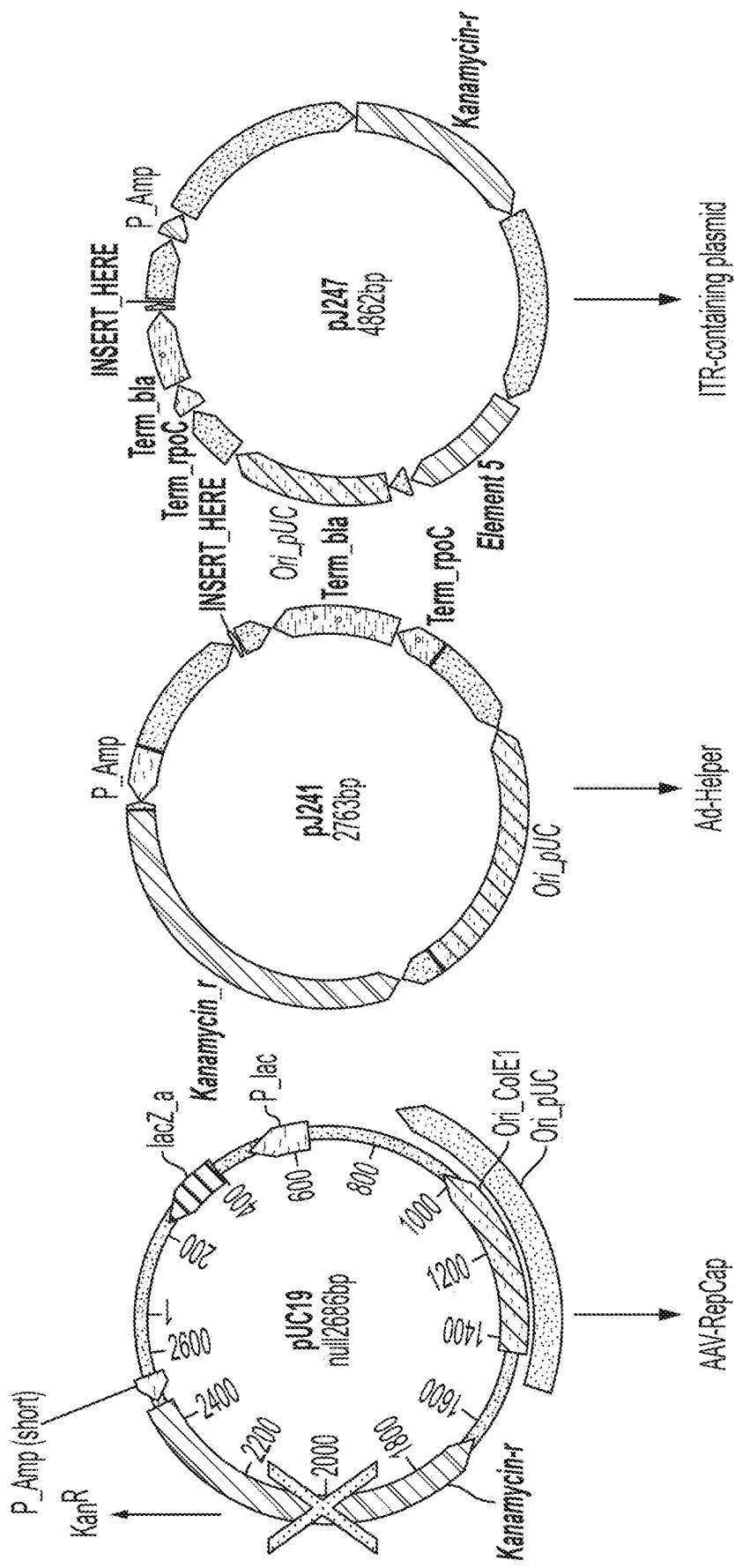


Figure 1

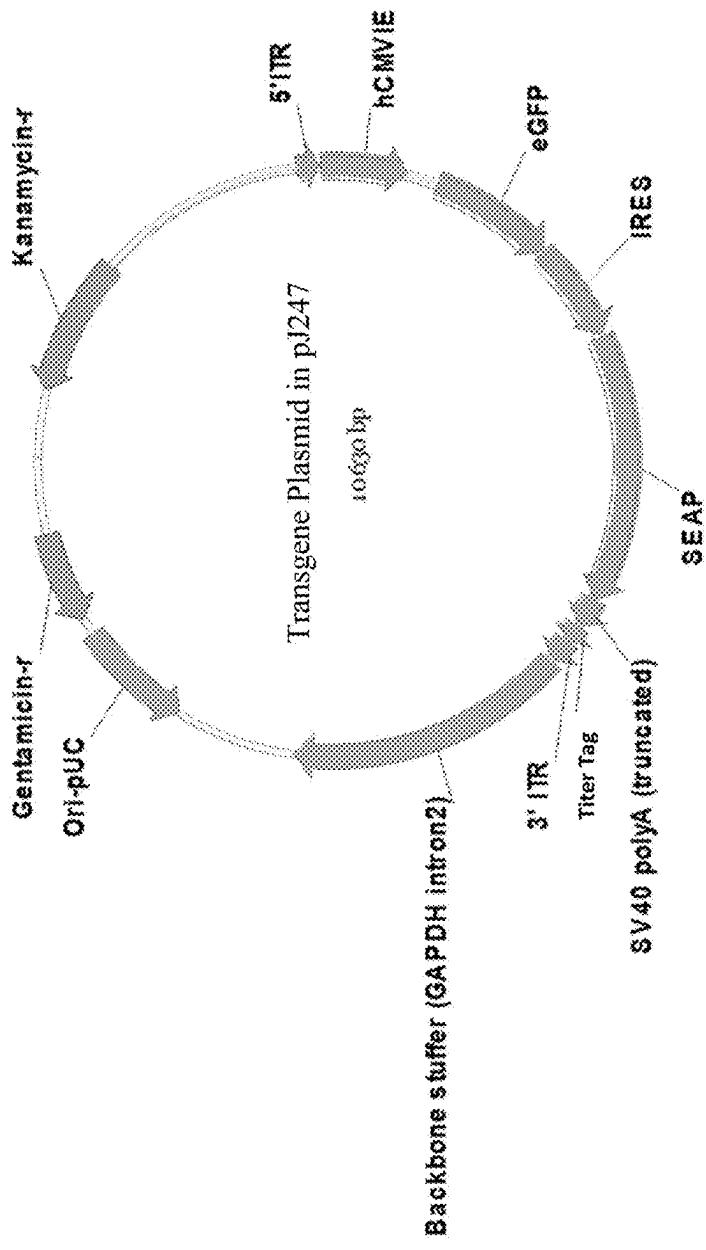


Figure 2

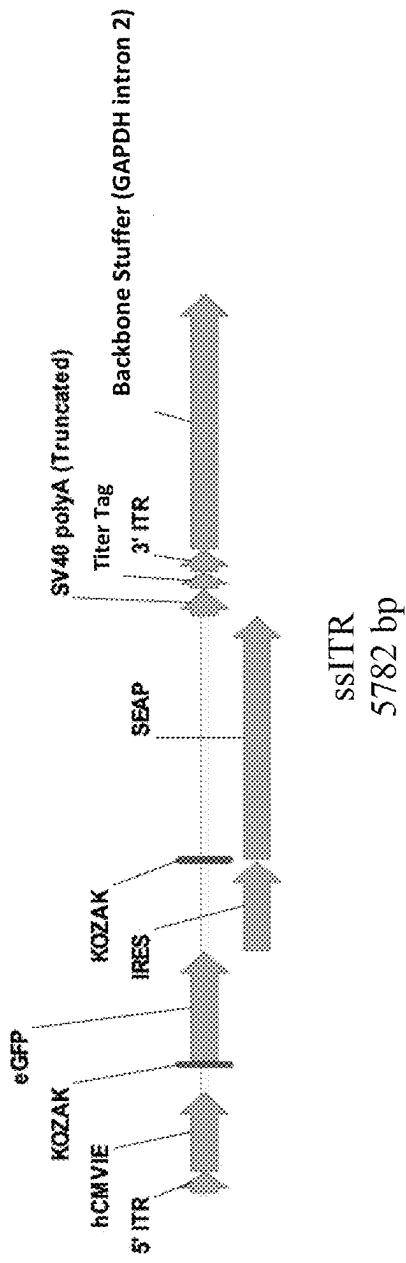


Figure 3A

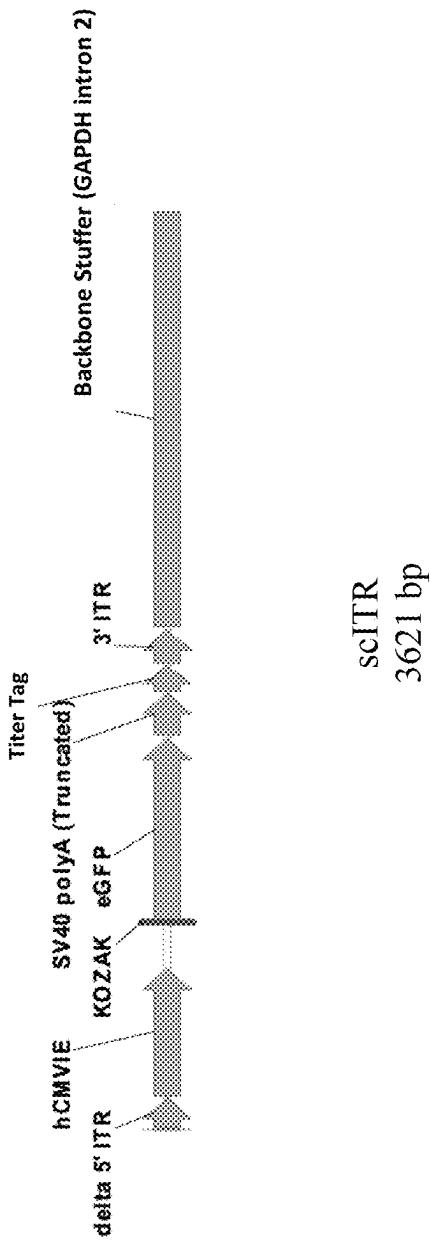


Figure 3B

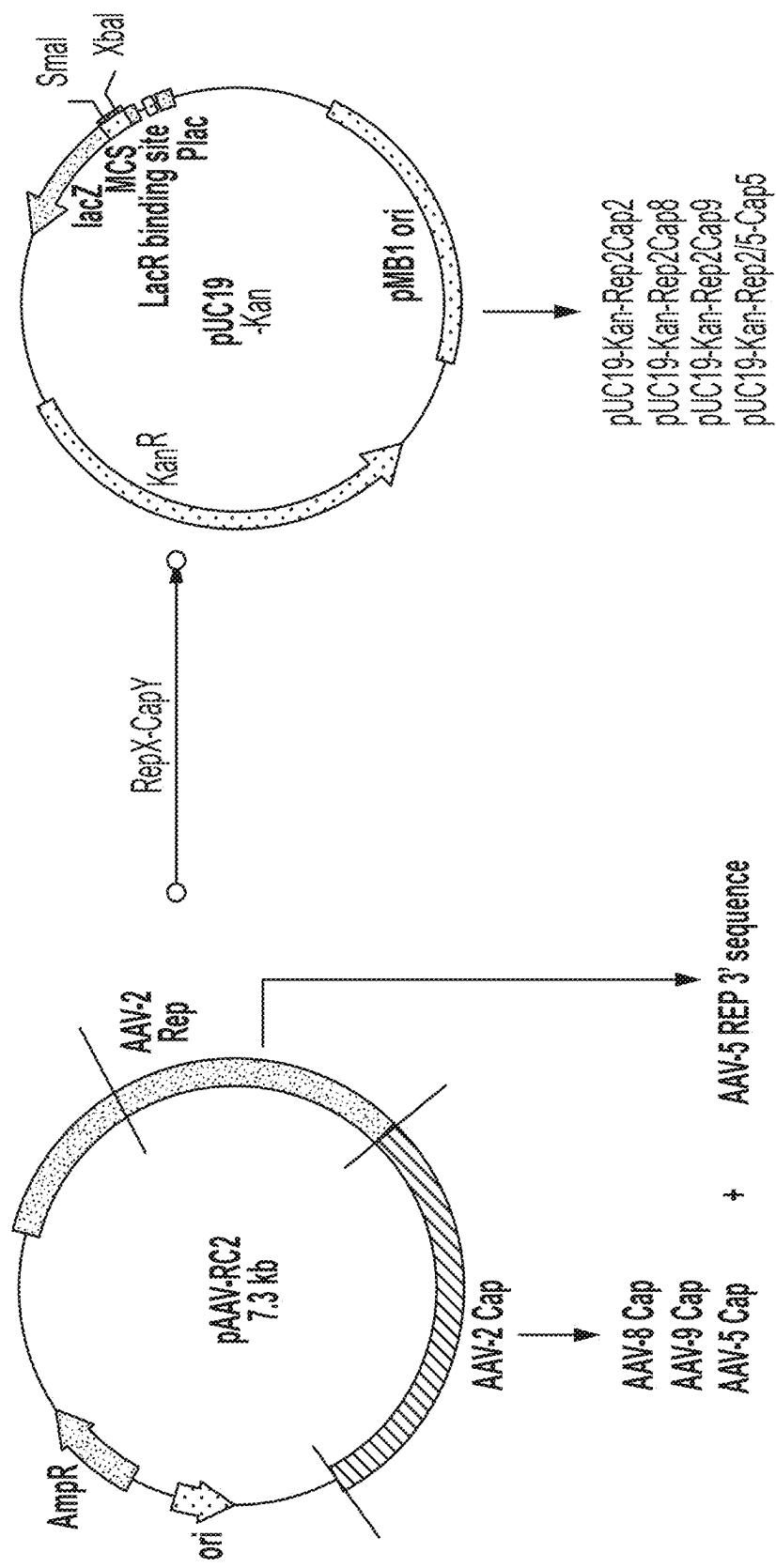


Figure 4A

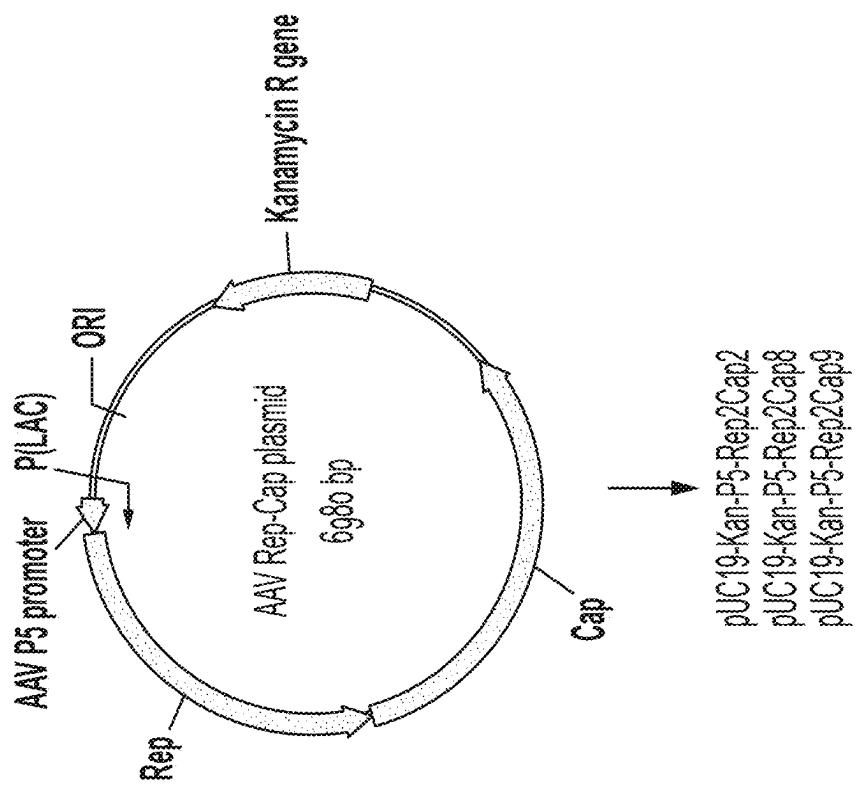


Figure 4B

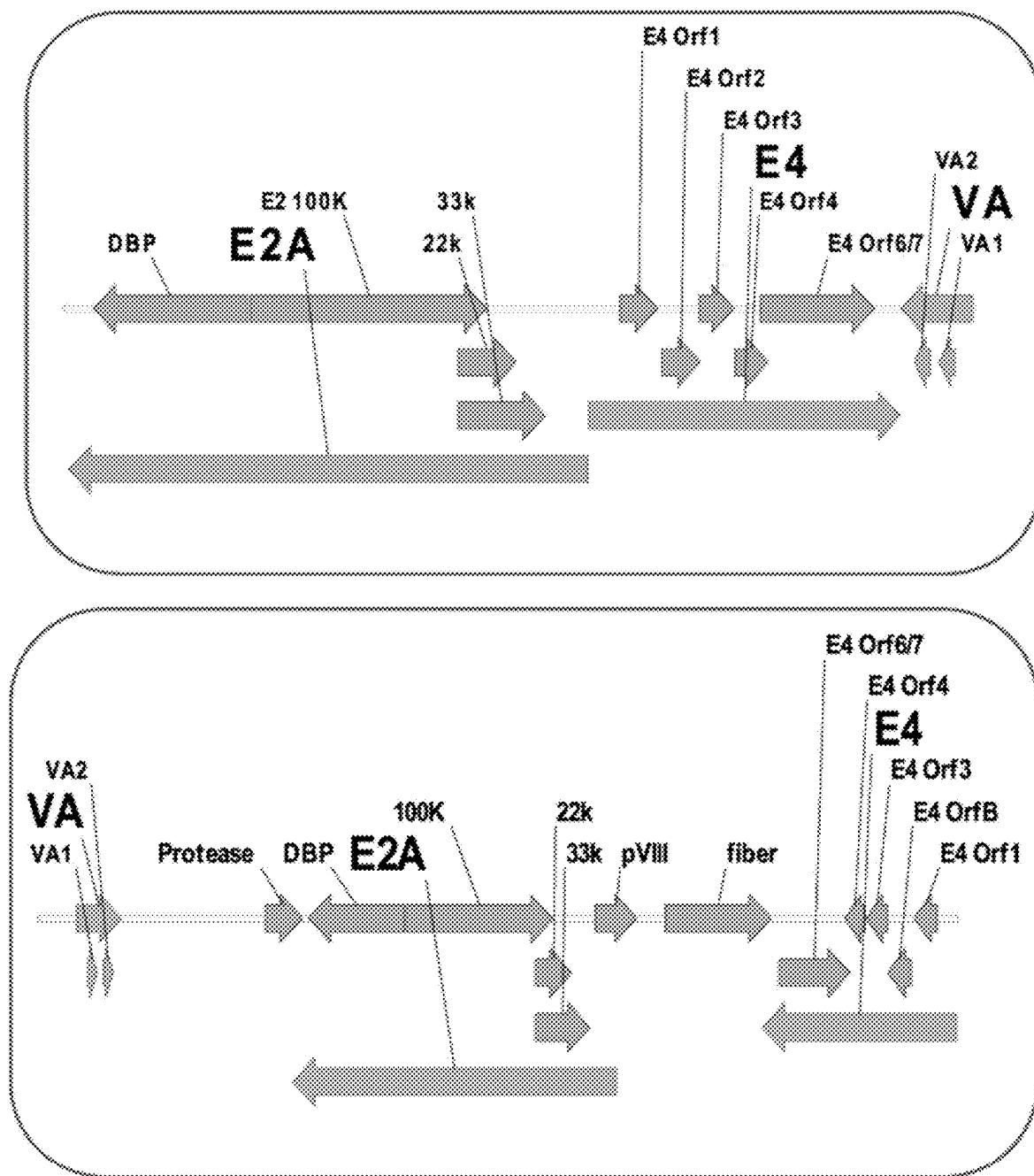


Figure 5

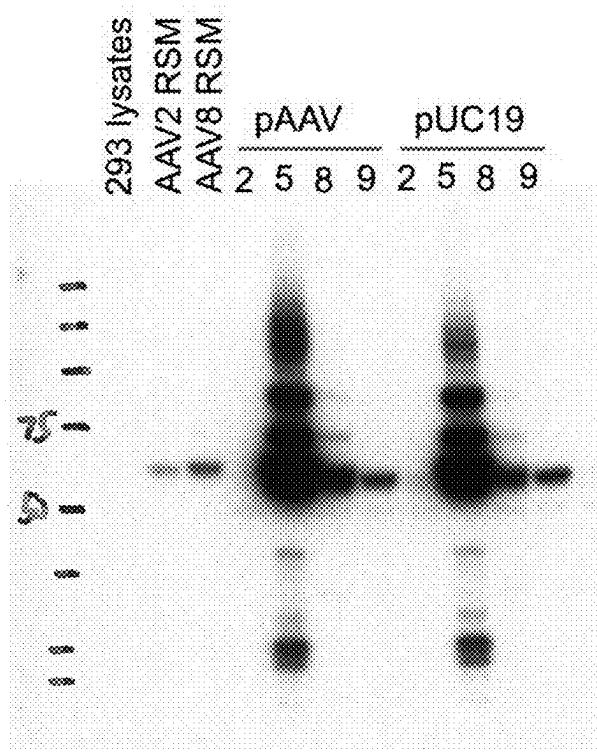


Figure 6

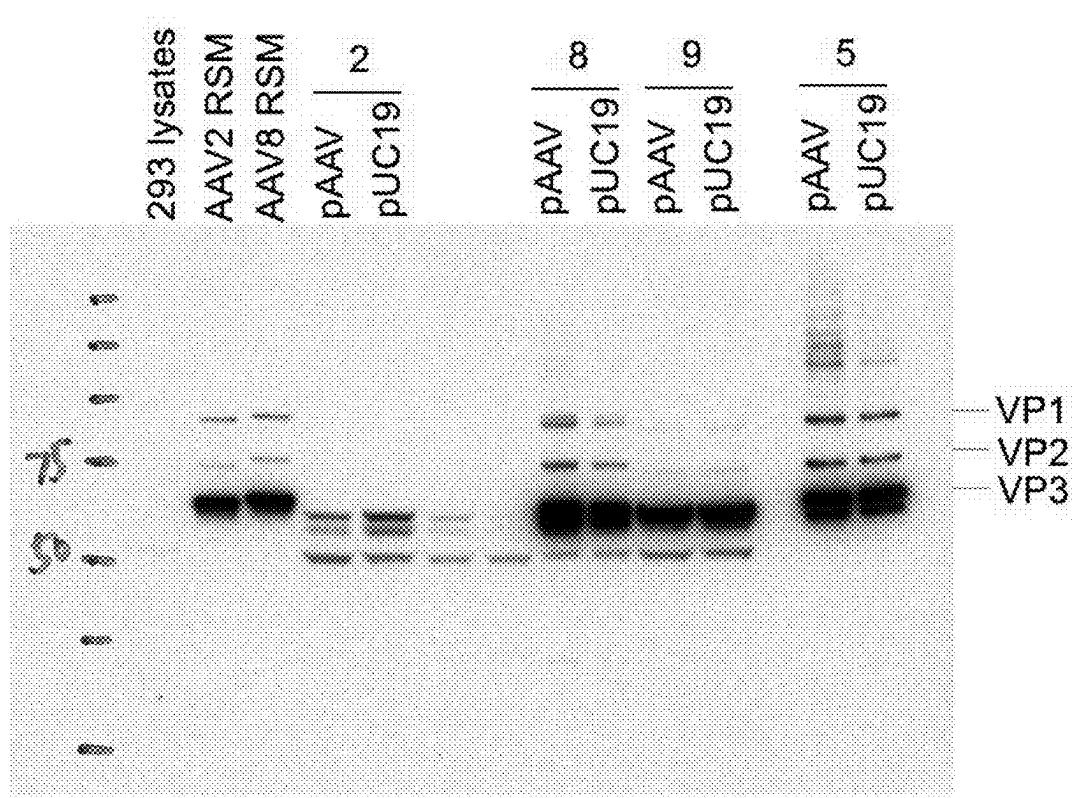


Figure 7

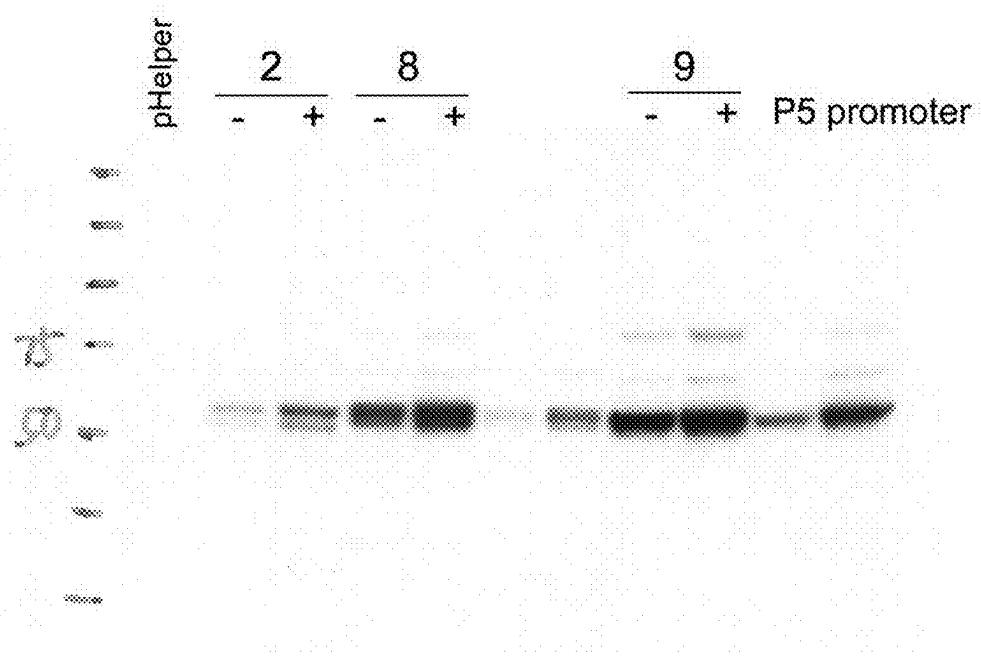


Figure 8

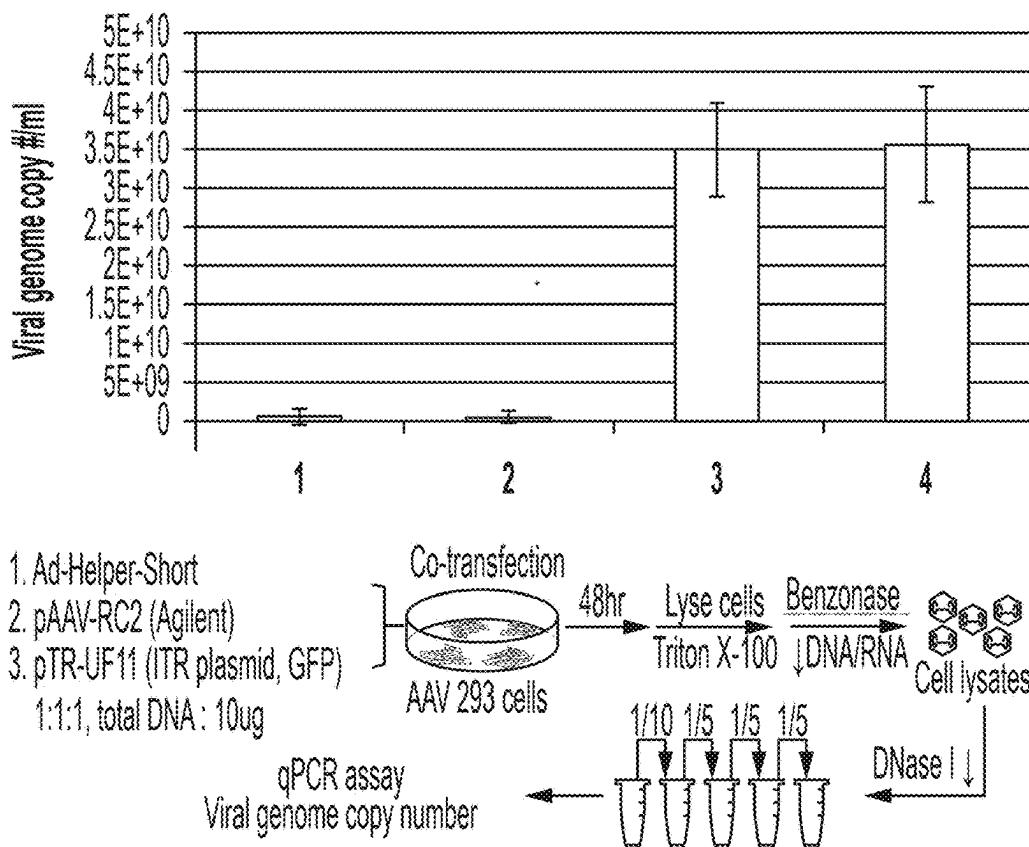


Figure 9

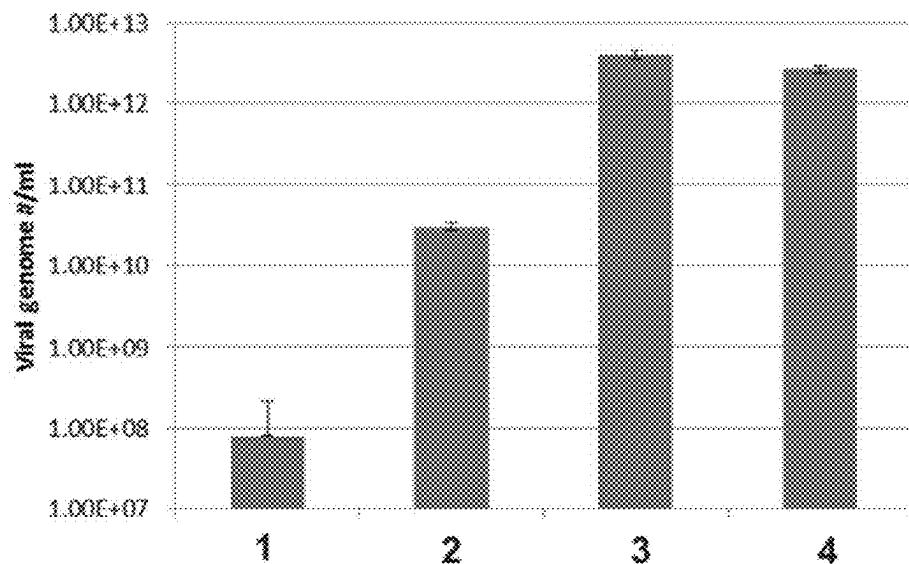


Figure 10

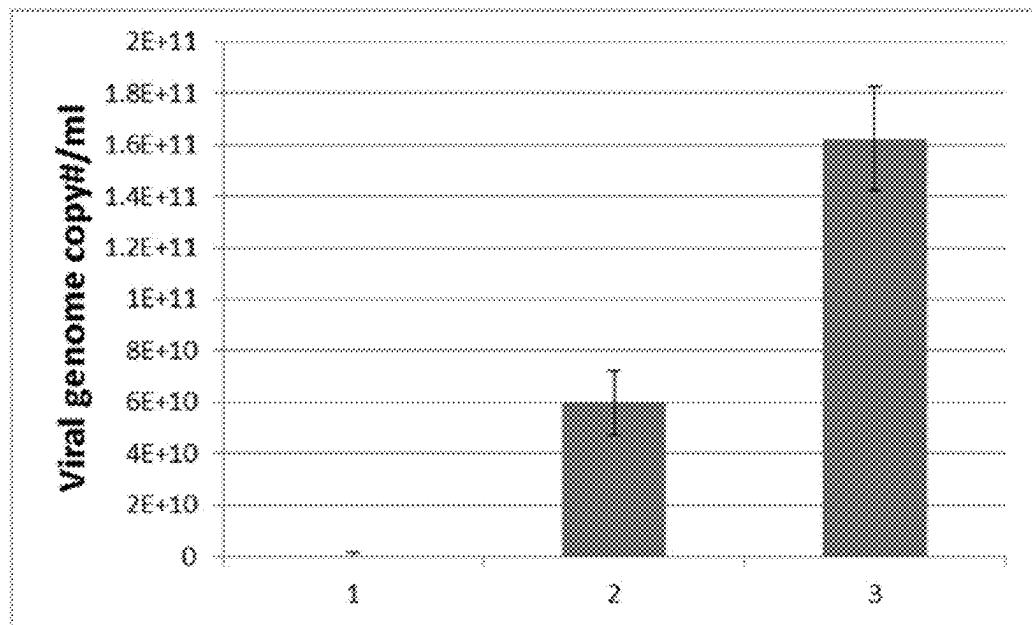
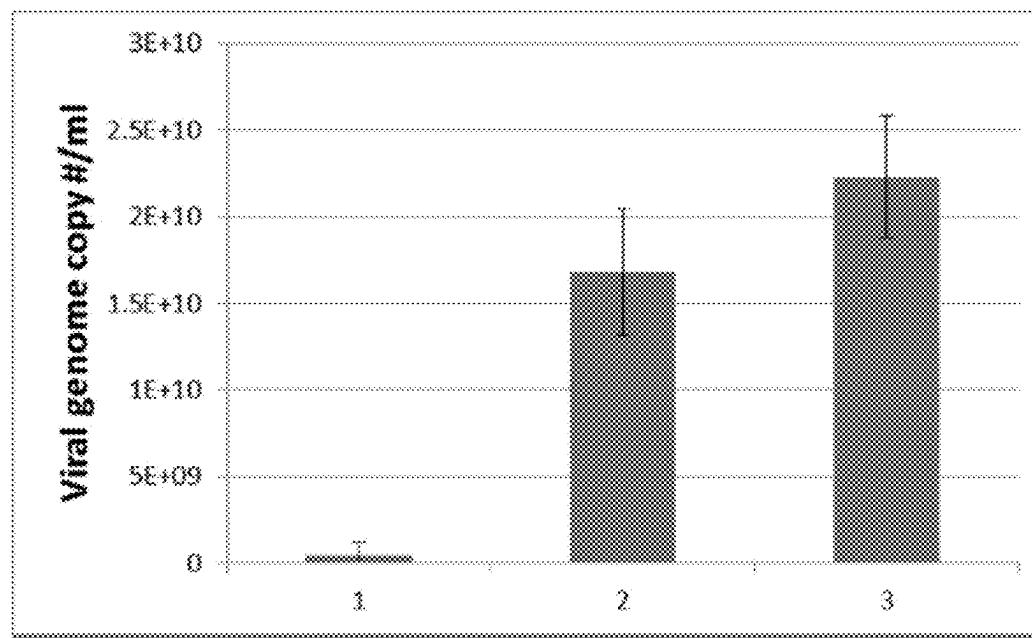
ssAAV**scAAV**

Figure 11

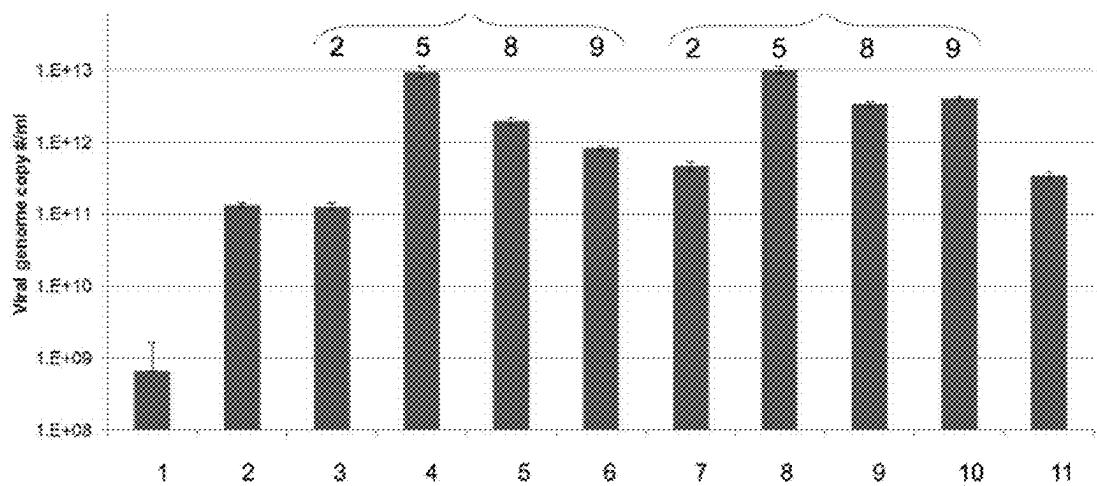


Figure 12

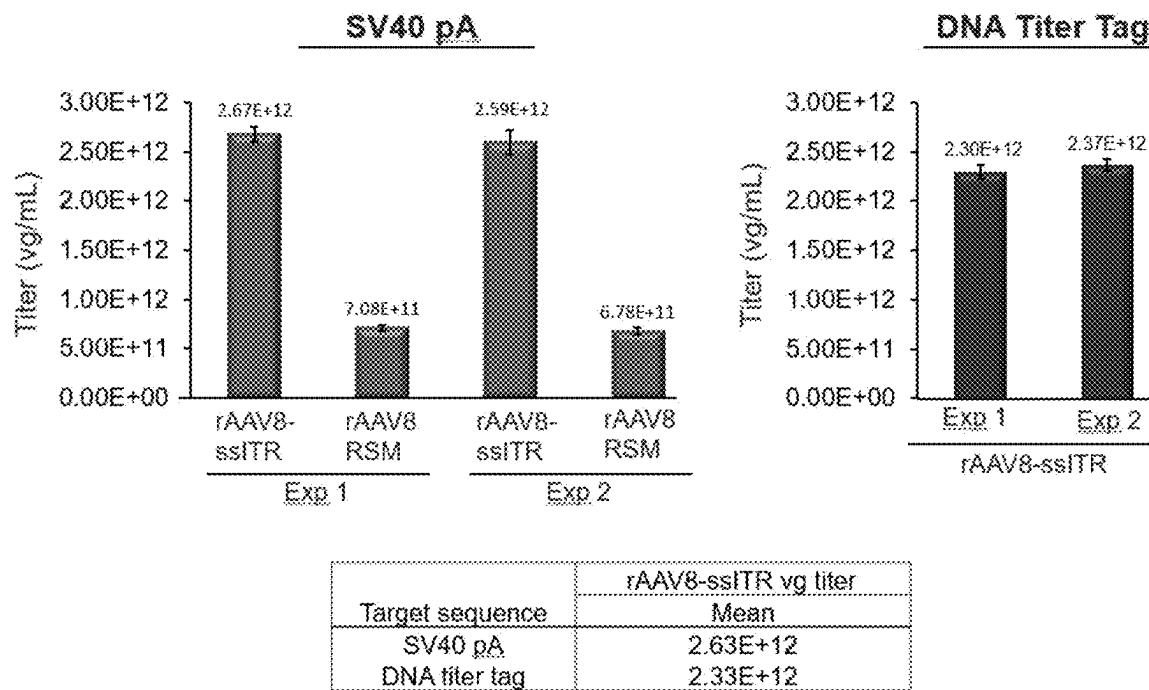


Figure 13A

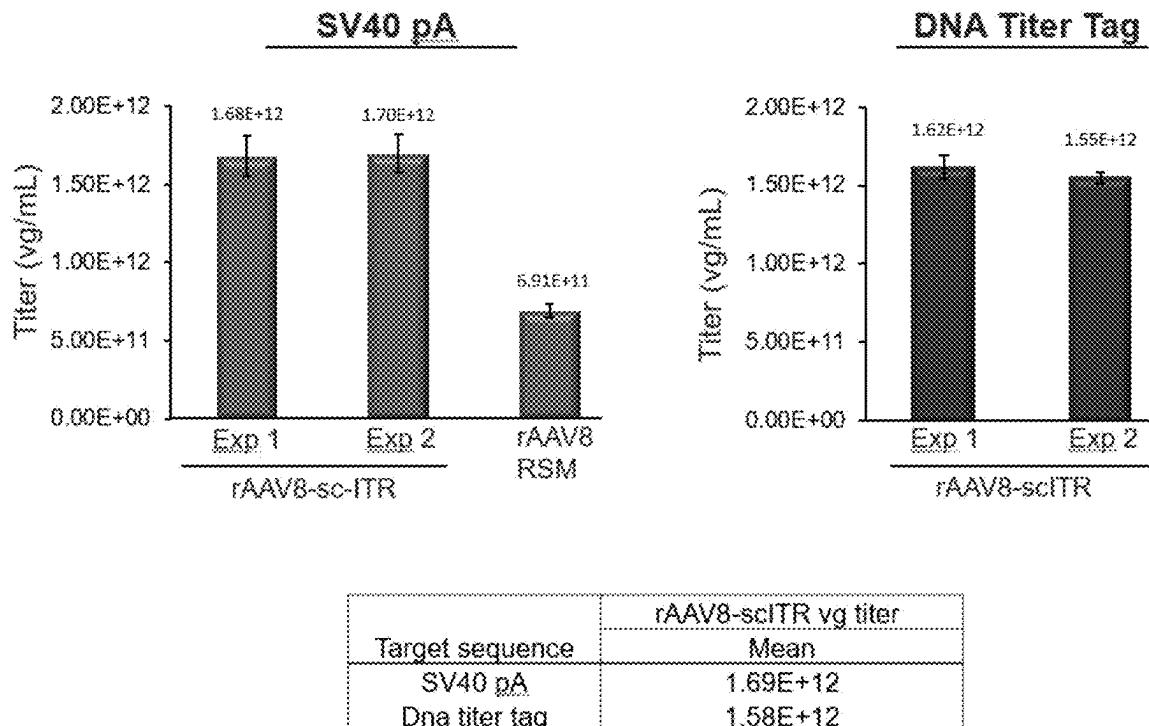


Figure 13B

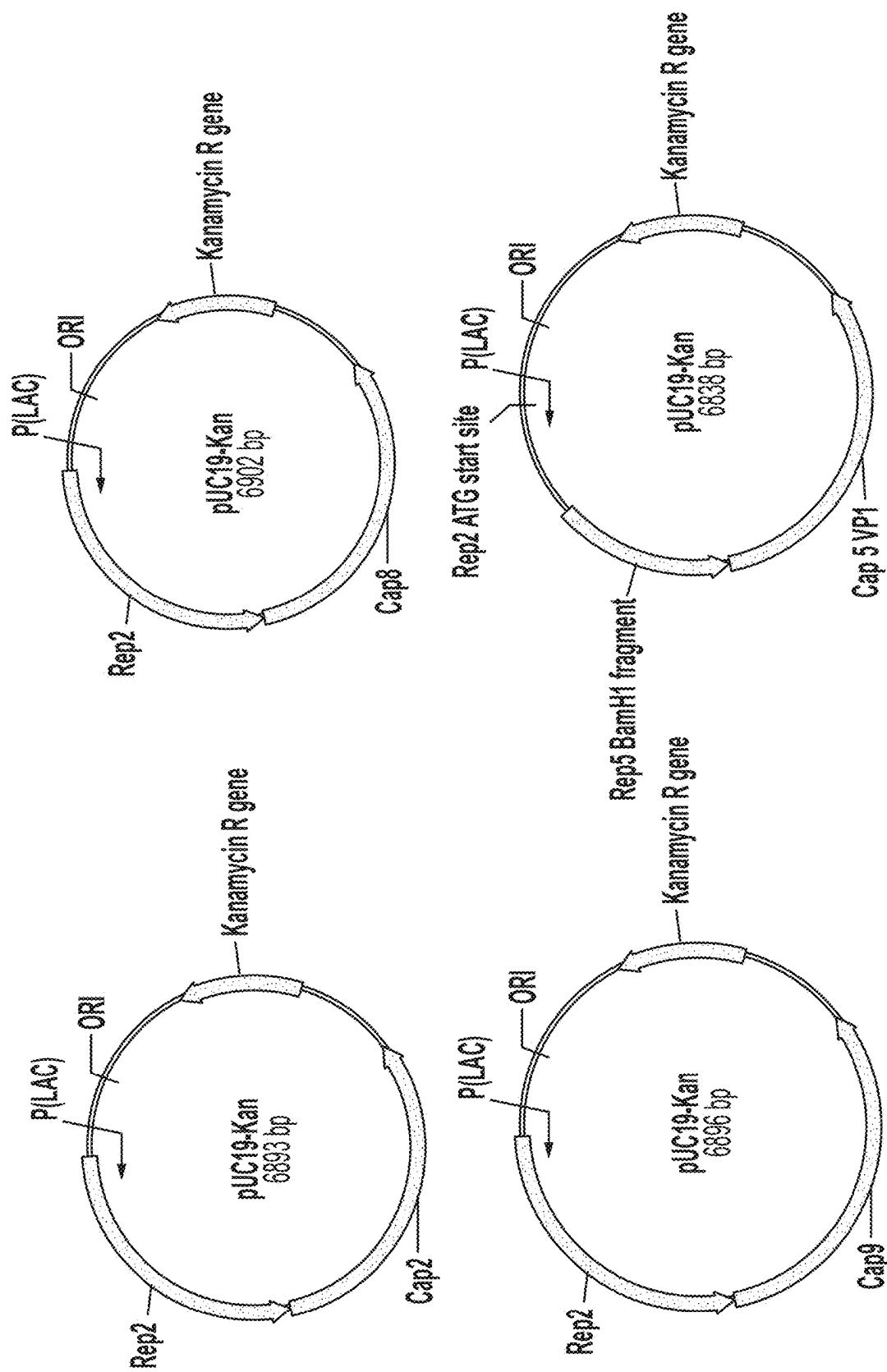


Figure 14A

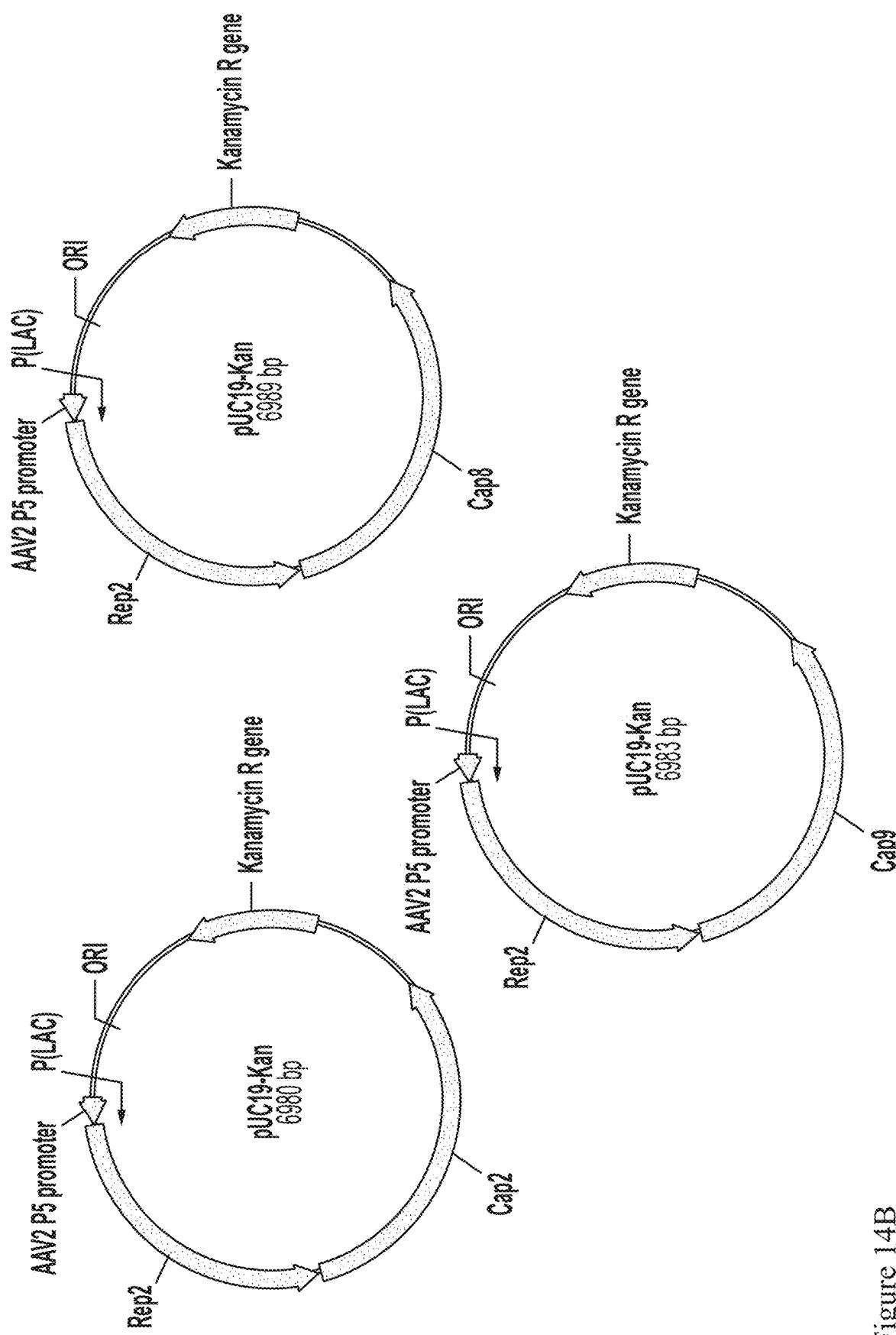


Figure 14B

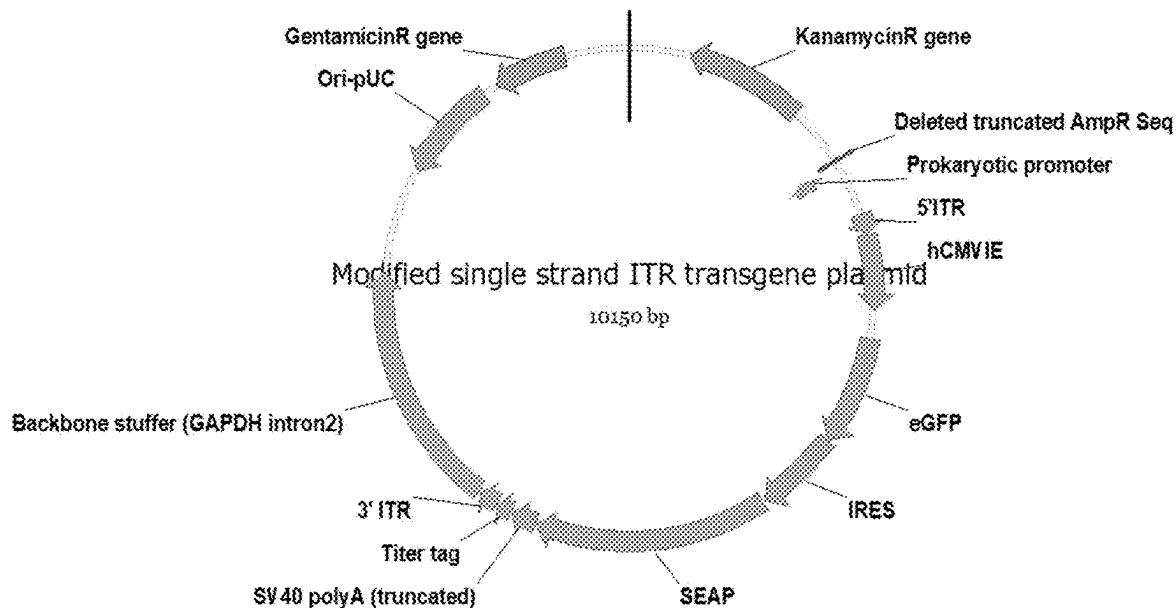


Figure 15A

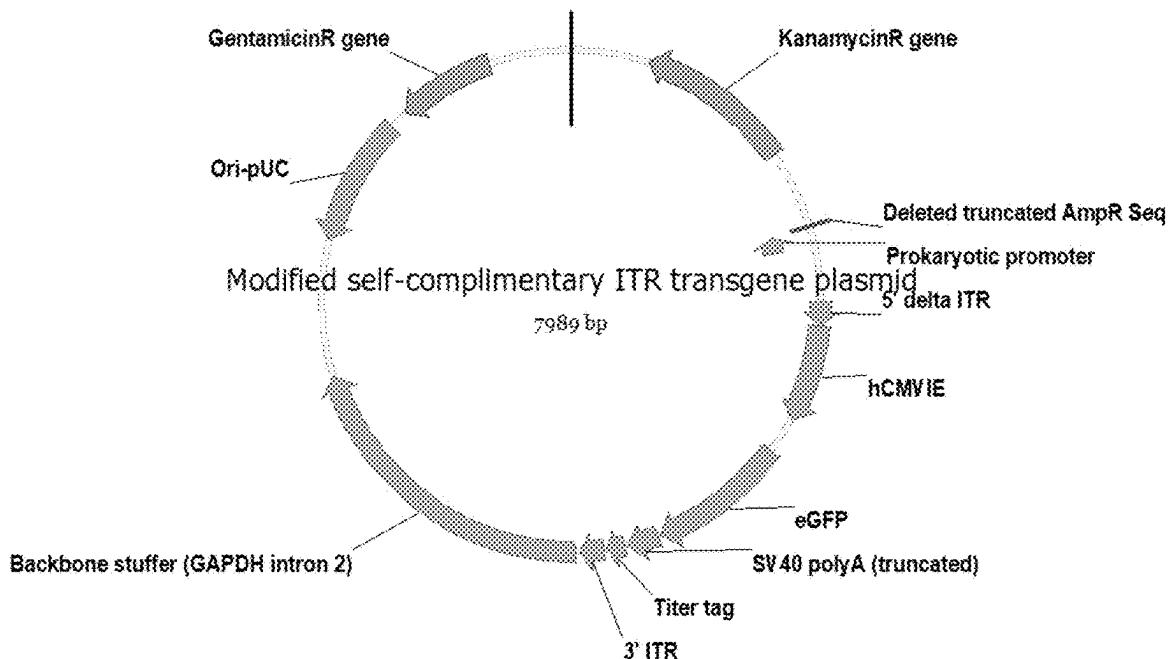


Figure 15B

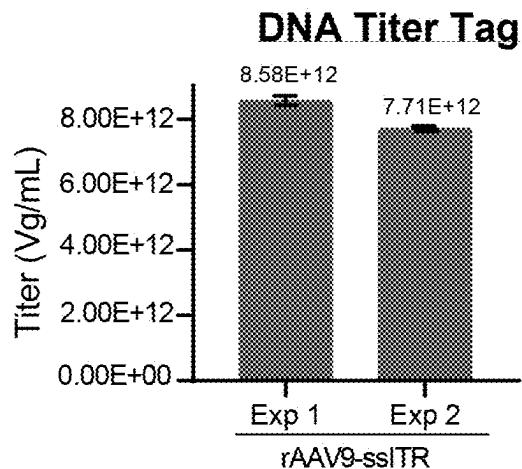


Figure 16A

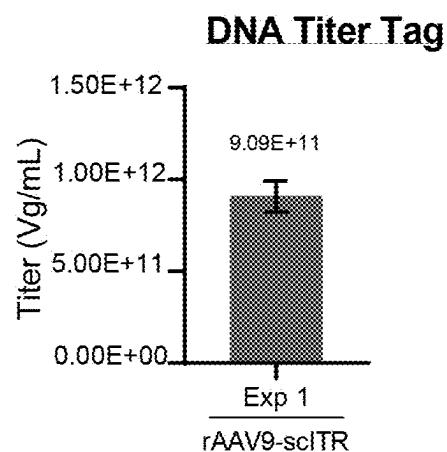


Figure 16B

1

AAV TRIPLE-PLASMID SYSTEM

CROSS-REFERENCE

This application is a 371 National Stage of PCT/US19/57916, filed Oct. 24, 2019, which claims priority U.S. Provisional Patent Application No. 62/750,603, filed Oct. 25, 2018, which are incorporated herein by reference in their entirety.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Oct. 22, 2019, is named 250478_001858_SL.txt and is 274,165 bytes in size.

BACKGROUND

Adeno-associated virus (AAV) is a DNA parvovirus that infects humans and various other animal species such as primates, bovine, feline, and canines. It belongs to the family Parvoviridae and is placed in the genus *Dependovirus*, because productive infection by AAV occurs only in the presence of a helper virus (e.g., adenovirus or herpes virus). This small non-enveloped virus contains a 4.6 kbases single stranded DNA genome that encodes sets of replication (Rep) and capsid (Cap) proteins. For example, Rep proteins (Rep78, Rep68, Rep52 and Rep40) are involved in replication, rescue and integration of the AAV genome, and Cap proteins (VP1, VP2 and VP3) provides structural function and form the virion capsid. Flanking the Rep and Cap open reading frames at the 5' and 3' ends are 145 bp inverted terminal repeats (ITRs). The ITRs function in cis as origins of nucleic acid replication and as packaging signals for the virus.

There are two stages to the AAV life cycle once infection has occurred: 1) the lytic stage and 2) the lysogenic stage. With the aid of the helper virus, the lytic stage begins. During this stage AAV commences productive infection resulting in genome replication, viral gene expression, and virion production. The case of the adenoviral helper, the adenoviral proteins that provide helper functions regarding AAV expression include E1a, E1b, E2a, E4, and VA RNA. The adenovirus helps regulate cellular gene expression by providing the proper milieu for AAV productive infection. See Daya and Berns *Clinical Microbiology Reviews* October 2008, p. 583-593.

AAV is a versatile virus that can be engineered for gene therapy. Recombinant Adeno-Associated Viral Vector (rAAV), which lacks viral genes in its DNA genome, used of gene therapy is primarily a protein-based nanoparticle engineered to cross the cell membrane in order to traffic and deliver its DNA cargo into the nucleus of a cell. rAAV DNA genome can form circular concatemers persisting as episomes in the nucleus of a transduced cell. As the rAAV DNA does not integrate into the host genome, which contributes to the long-term gene expression and durability, which is one of the reasons rAAV is ideal for gene therapy.

Recombinant forms of AAV (rAAV) have been developed as vectors by replacing all viral genes with a therapeutic transgene expression cassette, while retaining the only cis elements, the ITRs, which is required for vector packaging and DNA replication. See, e.g., U.S. Pat. Nos. 4,797,368; 5,153,414; 5,139,941; 5,252,479; and 5,354,678; and International Publication Nos. WO1991/018088; WO1993/

2

024641 and WO1994/13788. Early methods of rAAV production relied on a two-plasmid system comprising: 1) an AAV helper plasmid (generally encompassing AAV Rep and Cap coding regions, while lacking AAV ITRs so it cannot replicate or package itself) and 2) an ITR-containing plasmid (generally encompassing a selected transgene of interest bounded by AAV ITRs which provides for viral replication and packaging functions). Both the helper plasmid and the ITR-containing plasmid bearing the selected gene can be introduced into suitable cells for production by transient transfection. The transfected cell can then be infected with a helper virus, such as an adenovirus or herpes simplex virus, which transactivates the AAV promoters present on the helper plasmid that direct the transcription and translation of AAV Rep and Cap regions. Regarding the Ad helper virus, the E1a, E1b, E2a, E4, and VA RNA genes can supply the helper functions necessary for rAAV production. Infection of helper virus into producer cells to generate rAAV was effective in producing rAAV; however, a consequence is that it can also produce helper virus particles that can elicit immune responses from the host. In certain platforms, the viral helper genes necessary for AAV manufacturing can be stably transfected into the manufacturing cell line (e.g., HEK293 cells), thereby reducing the possibility of an anti-helper virus immune response by the host immune system coming from trace levels of residual helper virus.

More recently, a triple-plasmid transfection method has been developed. This method uses an AAV serotype-specific Rep and Cap plasmid as well as the transgene-containing plasmid but eliminated the use of helper virus infection by supplying the essential helper viral genes on a third plasmid (i.e., the viral coding sequences were removed or reduced), thus lowering the potential anti-helper virus immune response by the host immune system. Supplying the viral helper genes on the third plasmid greatly decreased helper viral production in the transfected cells, providing only rAAV. Multiplasmid transient transfection of adherent HEK293 cells is a commonly used method for rAAV production.

In a multiplasmid system, it is important to maintain an appropriate plasmid size. Thus, it may be important to add nucleic acid sequences (a.k.a “stuffer sequences”) to ensure that the plasmid is of an optimal size. For example, to discourage that the plasmid backbone of the ITR-containing plasmid is not packaged into the vector capsid, a stuffer sequence may need to be added such that the backbone is too large to be effectively packaged into the capsid. However, it is important that the stuffer sequence is “silent” and does not activate the immune system in the small chance that that plasmid does become packaged.

What is needed, therefore, is an improved triple-plasmid based system for producing rAAV. The plasmid system should provide improved transfection and lowered immunogenicity while still retaining optimum expression of the transgene. It is to such a plasmid system that embodiments of the present disclosure are directed.

BRIEF SUMMARY OF THE DISCLOSURE

As specified in the Background Section, there is a great need in the art to improve rAAV plasmid systems for rAAV-based gene therapies. The present disclosure satisfies this and other needs. Embodiments of the present disclosure relate generally to a plasmid system for the production of rAAV and more specifically to a triple-plasmid based system.

In one aspect, the invention is directed to a plasmid system for Recombinant Adeno-Associated Viral Vector (rAAV) production comprising: (i) a transgene-containing plasmid comprising at least one heterologous nucleic acid sequence flanked by a 5' and 3' AAV inverted terminal repeat (ITR) and a stuffer sequence outside of the ITRs; (ii) a plasmid comprising AAV replication (Rep) and capsid (Cap) gene sequences; and (iii) an adenovirus (Ad) helper plasmid.

In certain embodiments, the stuffer sequence increases the size of the transgene-containing plasmid backbone. In certain embodiments, the stuffer sequence increases the size of the transgene-containing plasmid backbone such that the transgene-containing plasmid backbone is discouraged from being packaged into an rAAV capsid. In certain embodiments, the plasmid backbone incorporation into the rAAV is below the limit of detection. In certain embodiments, the backbone of the transgene-containing plasmid is larger than a wild-type AAV genome following the addition of the stuffer sequence.

In certain embodiments, the stuffer sequence is devoid of enhancers, promoters, splicing regulators, noncoding RNAs, antisense sequences, coding sequences, or any combination thereof. In certain embodiments, the stuffer sequence is devoid of enhancers, promoters, splicing regulators, non-coding RNAs, antisense sequences, and coding sequences. In certain embodiments, the stuffer sequence comprises an inert intronic DNA sequence found in the human genome.

In certain embodiments, the stuffer sequence comprises a nucleic acid sequence of between 1000 and 5000 nucleotides in length or a nucleic acid sequence of between 1000 and 2000 nucleotides in length.

In certain embodiments, the stuffer sequence comprises GAPDH intron 2, fragment, or mutant thereof. In certain embodiments, the stuffer sequence comprises an inactivated gentamycin gene.

In certain embodiments, the stuffer sequence comprises a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 9. In certain embodiments, the stuffer sequences comprises SEQ ID NO: 9 or a fragment thereof. In certain embodiments, the fragment is between 800-1000 nucleotides long.

In certain embodiments, the stuffer sequence consists of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 9. In certain embodiments, the stuffer sequences consists of SEQ ID NO: 9 or a fragment thereof. In certain embodiments, the fragment is between 800-1000 nucleotides long.

In certain embodiments, the transgene-containing plasmid comprises a plasmid with a structure in the same order as FIG. 3A, wherein the eGFP and SEAP transgenes can be replaced with the at least one heterologous nucleic acid sequence.

In certain embodiments, the transgene-containing plasmid comprises a plasmid with a structure in the same order as FIG. 3B, wherein the eGFP transgene can be replaced with the at least one heterologous nucleic acid sequence.

In certain embodiments, the transgene-containing plasmid comprises nucleic acid sequences in the 5' to 3' direction of: a 5' ITR (e.g., SEQ ID NOs: 2 or 43), a promoter (e.g., SEQ ID NO: 4), at least one heterologous nucleic acid sequence, a polyA sequence (e.g., SEQ ID NO: 8), a 3' ITR (e.g., SEQ

ID NO: 3), and the stuffer sequence (e.g., SEQ ID NO: 9), wherein each nucleic acid sequence can be substituted with or encodes a corresponding functional fragment or derivative thereof or a sequence with at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity therewith.

In certain embodiments, the transgene-containing plasmid further comprises a DNA titer tag outside the expression cassette but between the 5' ITR and 3' ITR.

In certain embodiments, the transgene-containing plasmid further comprises a DNA titer tag i) upstream of the 3' ITR and downstream of a polyA sequence or ii) upstream of the 3' ITR and downstream of the at least one heterologous nucleic acid sequence; iii) or downstream of the 5' ITR and upstream of the at least one heterologous nucleic acid sequence; or iv) downstream of the 5' ITR and upstream of a promoter for the at least one heterologous nucleic acid sequence; or v) downstream of the 5' ITR and upstream of the 3' ITR.

In certain embodiments, the transgene-containing plasmid further comprises a DNA titer tag i) upstream of a 3' ITR (e.g., SEQ ID NO: 3) and downstream of a polyA sequence (e.g., SEQ ID NO: 8) or ii) upstream of a 3' ITR (e.g., SEQ ID NO: 3) and downstream of the at least one heterologous nucleic acid sequence; iii) or downstream of a 5' ITR (e.g., SEQ ID NOs: 2 or 43) and upstream of the at least one heterologous nucleic acid sequence; or iv) downstream of a 5' ITR (e.g., SEQ ID NOs: 2 or 43) and upstream of a promoter (e.g., SEQ ID NO: 4); or v) downstream of a 5' ITR (e.g., SEQ ID NOs: 2 or 43) and upstream of a 3' ITR (e.g., SEQ ID NO: 3).

In certain embodiments, the transgene-containing plasmid comprises nucleic acid sequences in the 5' to 3' direction of: a 5' ITR (e.g., SEQ ID NO: 2 or 43), a promoter (e.g., SEQ ID NO: 4), at least one heterologous nucleic acid sequence, a polyA sequence (e.g., SEQ ID NO: 8), a 3' ITR (e.g., SEQ ID NO: 3), and the stuffer sequence (e.g., SEQ ID NO: 9), wherein each nucleic acid sequence can be substituted with or encodes a corresponding functional fragment or derivative thereof or a sequence with at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity therewith.

In certain embodiments, the transgene-containing plasmid further comprises a DNA titer tag outside the expression cassette but between the 5' ITR and 3' ITR.

In certain embodiments, the transgene-containing plasmid further comprises a DNA titer tag i) upstream of the 3' ITR and downstream of a polyA sequence or ii) upstream of the 3' ITR and downstream of the at least one heterologous nucleic acid sequence; iii) or downstream of the 5' ITR and upstream of the at least one heterologous nucleic acid sequence; or iv) downstream of the 5' ITR and upstream of a promoter for the at least one heterologous nucleic acid sequence; or v) downstream of the 5' ITR and upstream of the 3' ITR.

In certain embodiments, the transgene-containing plasmid further comprises a DNA titer tag i) upstream of a 3' ITR (e.g., SEQ ID NO: 3) and downstream of a polyA sequence (e.g., SEQ ID NO: 8) or ii) upstream of a 3' ITR (e.g., SEQ ID NO: 3) and downstream of the at least one heterologous nucleic acid sequence; iii) or downstream of a 5' ITR (e.g., SEQ ID NOs: 2 or 43) and upstream of the at least one heterologous nucleic acid sequence; or iv) downstream of a 5' ITR (e.g., SEQ ID NOs: 2 or 43) and upstream of a

promoter (e.g., SEQ ID NO: 4); or v) downstream of a 5' ITR (e.g., SEQ ID NOS: 2 or 43) and upstream of a 3' ITR (e.g., SEQ ID NO: 3).

In certain embodiments, the AAV Rep gene sequence is from AAV serotype 2, 5, 8, 9, or hybrids thereof. In certain embodiments, the AAV Cap gene sequence is from AAV serotype 2, 5, 8, 9, or hybrids thereof. In certain embodiments, the plasmid comprising the Rep and Cap gene sequences further comprises a promoter. In certain embodiments, the promoter is an AAV promoter. In certain embodiments, the promoter is an AAV P5 promoter.

In certain embodiments, the Ad helper plasmid comprises one or more of Adenovirus genes selected from E1a, E1b, E2a, E4orf6, or VA RNA.

In certain embodiments, the Ad helper plasmid comprises nucleic acid sequences in the 5' to 3' direction of: SEQ ID NOS: 18, 17, 16, and 20, wherein each nucleic acid sequence can be substituted with a corresponding functional fragment or derivative thereof or a sequence with at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity therewith.

In certain embodiments, the Ad helper plasmid comprises nucleic acid sequences in the 5' to 3' direction of: SEQ ID NOS: 21, 16, 39, 40, 22, 23, and 20 wherein each nucleic acid sequence can be substituted with or encode a corresponding functional fragment or derivative thereof or a sequence with at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity therewith.

In certain embodiments, the Ad helper plasmid comprises a structure in the same order as either construct of FIG. 5.

In certain embodiments, the Ad helper plasmid comprises a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 14.

In certain embodiments, the Ad helper plasmid comprises a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 15.

In certain embodiments, the heterologous nucleic acid sequence is a heterologous gene of interest encoding a peptide, polypeptide, or protein. In certain embodiments, the peptide, polypeptide, or protein is an enzyme, antibody, MHC molecule, T-cell receptor, B-cell receptor, aptamer, avimer, receptor-binding ligand, targeting peptides, a therapeutic agent, or gene editing molecule. In certain embodiments, the heterologous nucleic acid sequence is a nucleic acid sequence such as an antisense, siRNA, shRNA, miRNA, EGSSs, gRNA, sgRNA, ribozyme, or aptamer.

In another aspect, the invention is directed to a host cell comprising any one of the plasmid systems described herein.

In another aspect, the invention is directed to a rAAV produced by any one of the plasmid systems described herein.

In another aspect, the invention is directed to a DNA titer tag allowing for universal vector titering, comprising a nucleic acid tag sequence from about 60 nucleotides to about 100 nucleotides long either upstream or downstream from a nucleic acid sequence of a heterologous nucleic acid sequence within a transgene-containing plasmid, wherein the nucleic acid tag sequence can be used in at least two

different transgene-containing plasmids to allow for universal vector genome titering between at least two different types of AAV vectors. In certain embodiments, the nucleic acid tag sequence is about 100 nucleotides long.

In certain embodiments, the nucleic acid tag sequence is upstream from a 3' ITR sequence of the transgene-containing plasmid but not within an expression cassette of the transgene-containing plasmid.

In certain embodiments, the nucleic acid tag sequence is downstream from a 5' ITR sequence of the transgene-containing plasmid but not within an expression cassette of the transgene-containing plasmid.

In certain embodiments, the DNA titer tag comprises any one of nucleic acid sequences of SEQ ID NOS: 61-70.

In another aspect, the invention is directed to a method for producing a rAAV comprising transducing a cell with the any one of the plasmid systems described herein and isolating the rAAV. In another aspect, the invention is directed to a rAAV produced by said method.

In another aspect, the invention is directed to a composition comprising the plasmid system of the invention.

In another aspect, the invention is directed to a pharmaceutical composition comprising the rAAV produced by the plasmid system of the invention.

In another aspect, the invention is directed to a method for delivering or transferring a nucleic acid sequence into a subject's cell, comprising administering the rAAV produced by the plasmid system of the invention to a subject thereby delivering the nucleic acid sequence into the cell. In certain embodiments, the subject's cell is in culture or is present in the subject.

In another aspect, the invention is directed to a method for treating or preventing a disease or disorder in a subject, comprising administering to a subject in need thereof a rAAV produced by the plasmid system of the invention.

In another aspect, the invention is directed to a host cell comprising contacting the host cell with a rAAV produced by the plasmid system of the invention.

These and other objects, features and advantages of the present disclosure will become more apparent upon reading the following specification in conjunction with the accompanying description, claims and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts an exemplary triple-plasmid system for the production of rAAV, in accordance with some embodiments of the present disclosure.

FIG. 2 depicts an exemplary transgene-containing plasmid for rAAV production incorporating eGFP and SEAP as transgenes, in accordance with some embodiments of the present disclosure.

FIGS. 3A-3B: shows exemplary gene constructs of transgene-containing plasmids for single-stranded (ss) (FIG. 3A) and self-complementary (sc) rAAV (FIG. 3B) production.

FIGS. 4A-4B: FIG. 4A depicts an exemplary AAV Rep-Cap plasmid incorporating different AAV Rep and Cap genes, in accordance with some embodiments of the present disclosure. FIG. 4B depicts an exemplary AAV Rep-Cap plasmid incorporating a promoter from AAV serotype 2.

FIG. 5 depicts exemplary Ad Helper Plasmids in short (top panel) and long (bottom panel) embodiments.

FIG. 6 is a Western blot showing expression levels of Cap proteins from different AAV serotypes from a plasmid according to the disclosure.

FIG. 7 is a Western blot showing expression levels of Cap proteins from different AAV serotypes from a plasmid according to the disclosure. A Monoclonal B1 clone was used for blot analysis.

FIG. 8 is a Western blot showing AAV P5-driven expression levels of Cap proteins from different AAV serotypes from a plasmid according to the disclosure. -: plasmid constructs without the P5 promoter; +: plasmid constructs with the P5 promoter. A Monoclonal B1 clone was used for blot analysis.

FIG. 9 shows the results of a qPCR assay of viral genome copy number using a short Ad Helper Plasmid according to the disclosure. 1: Negative Control 1 (pHelper+pAAV-RC2 (Agilent)); 2: Negative Control 2 (pHelper+pTRUF11); 3: Positive control (pHelper+pAAV-RC2+pTRUF11); 4: Short Ad Helper Test (Short-Helper (SEQ ID NO: 14)+pAAV-RC2+pTRUF11).

FIG. 10 shows the results of a qPCR assay of viral genome copy number using a long Ad Helper Plasmid according to the disclosure. 1: Negative Control 1 (pTRUF11+pAAV-RC2 (Rep2Cap2 (Agilent)); 2: Positive Control 2 (pHelper+pAAV-RC2+pTRUF11); 3: Short Ad Helper Test (Short-Helper (SEQ ID NO: 14)+pUC19-Rep2Cap8+pITRss (SEQ ID NO: 1)); 4: Long Ad Helper Test (Long-Helper (SEQ ID NO: 15)+pUC19-Rep2Cap8+pITRss (SEQ ID NO: 1)).

FIG. 11 shows the viral genome copy number per ml cell lysate for rAAV containing a single-stranded (top panel) or self-complementary (bottom panel) DNA genome produced using the corresponding transgene-containing plasmids for rAAV production. For the top panel: 1: Negative Control (pHelper+pAAV-RC2); 2: Positive Control (pHelper+pAAV-RC2+pTRUF11); 3: ssITR (pHelper+pAAV-RC2+ssITR) (SEQ ID NO: 1). For the bottom panel: 1: Negative Control (pHelper+pAAV-RC2); 2: Positive Control (pHelper+pAAV-RC2+pTRUF11); 3: scITR (pHelper+pAAV-RC2+scITR) (SEQ ID NO: 42).

FIG. 12 shows the viral genome copy number per ml cell lysate of multiple capsid serotypes for a triple-plasmid system according to the disclosure, along with positive and negative controls. 1: Negative Control (pHelper+pTRUF11); 2: Positive Control (pHelper+pTRUF11+pAAV-RC2); 3: pHelper+pTRUF11+pUC19-P5-Rep2Cap2 (SEQ ID NO: 31); 4: pHelper+pTRUF11+pUC19-Rep2/5Cap5 (SEQ ID NO: 24); 5: pHelper+pTRUF11+pUC19-P5-Rep2Cap8 (SEQ ID NO: 35); 6: pHelper+pTRUF11+pUC19-P5-Rep2Cap9 (SEQ ID NO: 37); 7: Short-Helper (SEQ ID NO: 14)+ssITR (SEQ ID NO: 1)+pUC19-P5-Rep2Cap2 (SEQ ID NO: 31); 8: Short-Helper (SEQ ID NO: 14)+ssITR (SEQ ID NO: 1)+pUC19-Rep2/5Cap5 (SEQ ID NO: 24); 9: Short-Helper (SEQ ID NO: 14)+ssITR (SEQ ID NO: 1)+pUC19-P5-Rep2Cap8 (SEQ ID NO: 35); 10: Short-Helper (SEQ ID NO: 14)+ssITR (SEQ ID NO: 1)+pUC19-P5-Rep2Cap9 (SEQ ID NO: 37); 11: Short-Helper (SEQ ID NO: 14)+ssITR (SEQ ID NO: 1)+pAAV-RC2.

FIG. 13A-13B shows that viral genome copy number per ml lysate for the single strand ITR (ssITR) transgene plasmid (FIG. 13A) and self complementary ITR (scITR) plasmid using both SV40 polyA and a 100 nucleotide long DNA titer tag for qPCR analysis.

FIGS. 14A-14B: FIG. 14A depicts an exemplary AAV Rep-Cap plasmid incorporating different AAV Rep and Cap genes, in accordance with some embodiments of the present disclosure. FIG. 14B depicts an exemplary AAV Rep-Cap plasmid incorporating different AAV Rep and Cap genes and incorporating the P5 promoter, in accordance with some embodiments of the present disclosure.

FIG. 15: shows exemplary gene constructs of transgene-containing plasmids for single-stranded (ss) (FIG. 15A) and self-complementary (sc) rAAV (FIG. 15B) production. Both modified plasmids were containing improved plasmid backbones with higher developability.

FIGS. 16A-16B: shows that viral genome copy number per ml lysate for the modified single strand ITR (ssITR) transgene plasmid (FIG. 16A) and modified self-complementary ITR (scITR) plasmid (FIG. 16B) using a 100 nucleotide long DNA titer tag for qPCR analysis.

DETAILED DESCRIPTION OF THE DISCLOSURE

As specified in the Background section, there is a great need in the art to identify technologies for rAAV production to generate rAAV-based gene therapies. The present disclosure satisfies this and other needs. Embodiments of the present disclosure relate generally to a rAAV production and more specifically to a triple-plasmid based system to produce rAAV.

To facilitate an understanding of the principles and features of the various embodiments of the disclosure, various illustrative embodiments are explained below. Although exemplary embodiments of the disclosure are explained in detail, it is to be understood that other embodiments are contemplated. Accordingly, it is not intended that the disclosure is limited in its scope to the details of construction and arrangement of components set forth in the following description or examples. The disclosure is capable of other embodiments and of being practiced or carried out in various ways. Also, in describing the exemplary embodiments, specific terminology will be resorted to for the sake of clarity.

It is intended that each term contemplates its broadest meaning as understood by those skilled in the art and includes all technical equivalents which operate in a similar manner to accomplish a similar purpose. It is to be understood that embodiments of the disclosed technology may be practiced without these specific details. In other instances, well-known methods, structures, and techniques have not been shown in detail in order not to obscure an understanding of this description. References to "one embodiment," "an embodiment," "example embodiment," "some embodiments," "certain embodiments," "various embodiments," etc., indicate that the embodiment(s) of the disclosed technology so described may include a particular feature, structure, or characteristic, but not every embodiment necessarily includes the particular feature, structure, or characteristic. Further, repeated use of the phrase "in one embodiment" does not necessarily refer to the same embodiment, although it may.

It must also be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the context clearly dictates otherwise. For example, reference to a component is intended also to include composition of a plurality of components. References to a composition containing "a" constituent is intended to include other constituents in addition to the one named. In other words, the terms "a," "an," and "the" do not denote a limitation of quantity, but rather denote the presence of "at least one" of the referenced item.

As used herein, the term "and/or" may mean "and," it may mean "or," it may mean "exclusive-or," it may mean "one,"

it may mean "some, but not all," it may mean "neither," and/or it may mean "both." The term "or" is intended to mean an inclusive "or."

Ranges may be expressed herein as from "about" or "approximately" or "substantially" one particular value and/or to "about" or "approximately" or "substantially" another particular value. When such a range is expressed, other exemplary embodiments include from the one particular value and/or to the other particular value. Further, the term "about" means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, "about" can mean within an acceptable standard deviation, per the practice in the art. Alternatively, "about" can mean a range of up to $\pm 20\%$, preferably up to $\pm 10\%$, more preferably up to $\pm 5\%$, and more preferably still up to $\pm 1\%$ of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, preferably within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated, the term "about" is implicit and in this context means within an acceptable error range for the particular value.

By "comprising" or "containing" or "including" is meant that at least the named compound, element, particle, or method step is present in the composition or article or method, but does not exclude the presence of other compounds, materials, particles, method steps, even if the other such compounds, material, particles, method steps have the same function as what is named.

Throughout this description, various components may be identified having specific values or parameters, however, these items are provided as exemplary embodiments. Indeed, the exemplary embodiments do not limit the various aspects and concepts of the present disclosure as many comparable parameters, sizes, ranges, and/or values may be implemented. The terms "first," "second," and the like, "primary," "secondary," and the like, do not denote any order, quantity, or importance, but rather are used to distinguish one element from another.

It is noted that terms like "specifically," "preferably," "typically," "generally," and "often" are not utilized herein to limit the scope of the claimed disclosure or to imply that certain features are critical, essential, or even important to the structure or function of the claimed disclosure. Rather, these terms are merely intended to highlight alternative or additional features that may or may not be utilized in a particular embodiment of the present disclosure. It is also noted that terms like "substantially" and "about" are utilized herein to represent the inherent degree of uncertainty that may be attributed to any quantitative comparison, value, measurement, or other representation.

The dimensions and values disclosed herein are not to be understood as being strictly limited to the exact numerical values recited. Instead, unless otherwise specified, each such dimension is intended to mean both the recited value and a functionally equivalent range surrounding that value. For example, a dimension disclosed as "50 mm" is intended to mean "about 50 mm".

It is also to be understood that the mention of one or more method steps does not preclude the presence of additional method steps or intervening method steps between those steps expressly identified. Similarly, it is also to be understood that the mention of one or more components in a composition does not preclude the presence of additional components than those expressly identified.

As used herein, the terms "subject", "patient", "individual", and "animal" are used interchangeably herein and refer to mammals, including, without limitation, human and veterinary animals (e.g., cats, dogs, cows, horses, sheep, pigs, etc.) and experimental animal models. In a preferred embodiment, the subject is a human.

As used herein, the term "gene therapy" includes any therapeutic approach of providing a nucleic acid encoding a therapeutic gene (e.g., a Factor VIII/IX/X) to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with a disease or condition. The term encompasses administering any compound, drug, procedure, or regimen comprising a nucleic acid encoding a therapeutic gene, including any modified form of the gene (e.g., a Factor VIII/IX/X variant), for maintaining or improving the health of an individual with the disease or condition. One skilled in the art will appreciate that either the course of gene therapy or the dose of a genetic therapeutic agent can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

As used herein the term "therapeutically effective" applied to dose or amount refers to that quantity of a compound or pharmaceutical composition that when administered to a subject for treating (e.g., preventing or ameliorating) a state, disorder or condition, is sufficient to affect such treatment. For example, a therapeutically effective amount of a drug useful for treating hemophilia can be the amount that is capable of preventing or relieving one or more symptoms associated with hemophilia. The "therapeutically effective amount" will vary depending on the compound or bacteria or analogues administered as well as the disease and its severity and the age, weight, physical condition and responsiveness of the mammal to be treated. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques (see, e.g., Lieberman, Pharmaceutical Dosage Forms (vols. 1-3, 1992); Lloyd, The Art, Science and Technology of Pharmaceutical Compounding (1999); Pickar, Dosage Calculations (1999); and Remington: The Science and Practice of Pharmacy, 20th Edition, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

As used herein, the term "vector" refers to any vehicle used to transfer a nucleic acid (e.g., encoding a gene therapy construct) into a host cell. In some embodiments, a vector includes a replicon, which functions to replicate the vehicle, along with the target nucleic acid. In some embodiments, a vector is a viral particle for introducing a target nucleic acid (e.g., a codon-altered polynucleotide encoding a therapeutic gene or therapeutic gene variant). Many modified eukaryotic viruses useful for gene therapy are known in the art. For example, adeno-associated viruses (AAVs) are particularly well suited for use in human gene therapy because humans are a natural host for the virus, the native viruses are not known to contribute to any diseases, and the viruses elicit a mild immune response. "Recombinant AAV" (rAAV) and "AAV" are used interchangeably throughout the application.

The term "plasmid" refers to an extrachromosomal circular DNA capable of autonomous replication in a given bacterial cell. Exemplary plasmids include but are not limited to those derived from pBR322, pUC, pUC19, pUC57, pJ241, or pJ247, pBluescript, pREP4, pCEP4, pCI, and p Poly (Lathe et al., Gene 57 (1987), 193-201). Plasmids can also be engineered by standard molecular biology techniques (Sambrook et al., Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), N.Y.). It may also comprise a selection gene in order to select or to

11

identify the transfected cells (e.g., by complementation of a cell auxotrophy or by antibiotic resistance), stabilizing elements (e.g., cer sequence) or integrative elements (e.g., LTR viral sequences and transposons).

As used herein, the term "plasmid backbone" refers to a sequence of DNA that typically contains an origin of replication (e.g., SEQ ID NOS: 20 and 26), and an antibiotic selection gene, which are necessary for the specific growth of only the host that is transformed with the proper plasmid. In certain embodiments, these elements are not intended to be packaged in the rAAV capsid.

As used herein, the term "gene" refers to the segment of a DNA molecule that codes for a polypeptide chain (e.g., the coding region). In some embodiments, a gene is positioned by regions immediately preceding, following, and/or intervening the coding region that are involved in producing the polypeptide chain (e.g., regulatory elements such as a promoter, enhancer, polyadenylation sequence, 5'-untranslated region, 3'-untranslated region, or intron).

As used herein, the term "regulatory elements" refers to nucleic acid sequences, such as promoters, enhancers, terminators, polyadenylation sequences, introns, etc . . . , that provide for the expression of a coding sequence in a cell.

As used herein, the term "promoter element" refers to a nucleic acid sequence that assists with controlling expression of a coding sequence. Generally, promoter elements are located 5' of the translation start site of a gene. However, in certain embodiments, a promoter element may be located within an intron sequence, or 3' of the coding sequence. In some embodiments, a promoter useful for gene therapy is derived from the native gene of the target protein. In some embodiments, a promoter useful for gene therapy is specific for expression in a particular cell or tissue of the target organism (e.g., a liver-specific promoter) (Wu Z et al. Molecular Therapy 16(2):280-9. Choi V W et al. Molecular Therapy Methods & Clinical Development 2015. 2:15022), both of which are incorporated herein in their entirety for all intended purposes. In yet other embodiments, one of a plurality of well characterized promoter elements is used in gene therapy described herein.

Non-limiting examples of well-characterized promoter elements include the CMV early promoter (e.g., hCMVie (SEQ ID NO: 4)), the 3-actin promoter, and the methyl CpG binding protein 2 (MeCP2) promoter. In some embodiments, the promoter is a constitutive promoter, which drives substantially constant expression of the target protein. In other embodiments, the promoter is an inducible promoter, which drives expression of the target protein in response to a particular stimulus (e.g., exposure to a particular treatment or agent). For a review of designing promoters for AAV-mediated gene therapy, see Gray et al. (Human Gene Therapy 22:1143-53 (2011)), the contents of which are expressly incorporated by reference in their entirety for all purposes.

As used herein, the term "transgene" broadly refers to any nucleic acid that is introduced into the genome of an animal, including but not limited to genes or nucleic acid having sequences which are perhaps not normally present in the genome, genes which are present but not normally transcribed and translated ("expressed") in a given genome, or any other gene or nucleic acid which one desires to introduce into the genome. This may include genes which may normally be present in the non-transgenic genome, but which one desires to have altered in expression, or which one desires to introduce in a non-mutated form or an altered or variant form. The transgene may be specifically targeted to a defined genetic locus, may be randomly integrated within

12

a chromosome, or it may be extrachromosomally replicating DNA. A transgene may include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid. A transgene can be as few as a couple of nucleotides long, but is preferably at least about 50, 100, 150, 200, 250, 300, 350, 400, or 500 nucleotides long or even longer and can be, e.g., an entire viral genome. A transgene can be coding or non-coding sequences, or a combination thereof. A transgene usually comprises a regulatory element that is capable of driving the expression of one or more transgenes under appropriate conditions.

As used herein, the term "heterologous" as it relates to nucleic acid sequences, such as coding sequences and/or control sequences, denotes sequences that are not normally joined together and/or are not normally associated with a particular cell. Thus, a "heterologous" nucleic acid sequence means that the nucleic acid sequence is from an organism other than AAV or is synthetically derived. In certain embodiments, the heterologous nucleic acid sequence (e.g., a heterologous gene of interest) can encode a polypeptide such as, but not limited to, a clotting factor, an enzyme, an antibody or other polypeptide of interest. In certain embodiments, the heterologous nucleic acid sequence can encode an RNA having a structural or therapeutic function such as, but not limited to, an antisense, siRNA, shRNA, miRNA, EGSS, gRNA, sgRNA, ribozyme, or aptamer. Similarly, a cell transformed with a construct which is not normally present in the cell would be considered heterologous for purposes of this invention.

"Operably-linked" refers to the association of two or more nucleic acid sequence elements that are physically linked so that the function of one of the sequences is affected by another. For example, a regulatory DNA sequence is said to be "operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (i.e., that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

As used herein, the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form and complements thereof. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, and peptide-nucleic acids (PNAs).

The term "amino acid" refers to naturally occurring and non-natural amino acids, including amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids include those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Naturally occurring amino acids can include, e.g., D- and L-amino acids. The amino acids used herein can also include non-natural amino acids. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., any carbon that is bound to a

hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, or methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally occurring amino acid. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

The term "derivative" as used herein refers to a nucleic acid, peptide, or protein or a variant or analog thereof comprising one or more mutations and/or chemical modifications as compared to a corresponding full-length wild type nucleic acid, peptide or protein. Non-limiting examples of chemical modifications involving nucleic acids include, for example, modifications to the base moiety, sugar moiety, phosphate moiety, phosphate-sugar backbone, or a combination thereof.

The nucleic acid sequences that encode mutant gene constructs that may be useful with the plasmid system described herein may be identical to a wildtype (i.e., unmutated) sequence or may be a different coding sequence, which sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptides as the wildtype coding sequence. One of ordinary skill in the art will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each variation of a nucleic acid which encodes a same polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual gene therapy constructs.

As to amino acid sequences, one of ordinary skill in the art will recognize that individual substitutions, deletions or additions to a nucleic acid or peptide sequence that alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure. Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. Dependent on the functionality of the particular amino acid, e.g., catalytic, structural, or sterically important amino acids, different groupings of amino acid may be considered conservative substitutions for each other.

The terms "identical" or percent (%) "identity," in the context of two or more nucleic acids or peptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST

2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection.

As is known in the art, a number of different programs 5 may be used to identify whether a protein (or nucleic acid as discussed below) has sequence identity or similarity to a known sequence. Sequence identity and/or similarity is determined using standard techniques known in the art, including, but not limited to, the local sequence identity 10 algorithm of Smith & Waterman, *Adv. Appl. Math.*, 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.*, 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. U.S.A.*, 85:2444 (1988), by computerized 15 implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.), the Best Fit sequence program described by Devereux et al., *Nucl. Acid Res.*, 12:387-395 (1984), preferably using the default settings, or by inspection. Preferably, percent identity is calculated by FastDB based upon the 20 following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, 25 "Current Methods in Sequence Comparison and Analysis," Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc, all of which are incorporated by reference.

In accordance with the present disclosure there may be 30 employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring 35 Harbor Laboratory Press, Cold Spring Harbor, New York (herein "Sambrook et al., 1989"); *DNA Cloning: A Practical Approach*, Volumes I and II (D. N. Glover ed. 1985); Oligonucleotide Synthesis (M. J. Gait ed. 1984); *Nucleic Acid Hybridization* (B. D. Hames & S. J. Higgins eds. 40 (1985); *Transcription and Translation* (B. D. Hames & S. J. Higgins, eds. (1984); *Animal Cell Culture* (R. I. Freshney, ed. (1986); *Immobilized Cells and Enzymes* (IRL Press, (1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); F. M. Ausubel et al. (eds.), *Current Protocols in 45 Molecular Biology*, John Wiley & Sons, Inc. (1994); among others.

Plasmid Systems of the Disclosure

In one aspect, the disclosure provides a triple-plasmid 50 system for engineering and producing Recombinant Adeno Associated Viral Vector (rAAV). In certain embodiments, the three plasmid backbones are all the same. In certain embodiments, at least one of the three plasmid backbones are different. In certain embodiments, all three plasmid backbones are different. In certain embodiments, all three 55 plasmid backbones are different to prevent recombination occurring that can lead to the reconstruction of the complete AAV genome. In certain embodiments, the three plasmids comprise plasmid backbones based on, for example and without limitation, pUC19, pBR322, pUC57, pJ241, or pJ247. In certain embodiments, the three plasmids comprise 60 plasmid backbones based on pUC19, pJ241, and pJ247.

In certain embodiments, one plasmid serves as the trans-gene-containing plasmid for rAAV production construct, a second plasmid serves as the AAV Rep-Cap construct, and a third plasmid serves as the Adenovirus (Ad) Helper construct. Exemplary plasmids of each type are shown in FIG. 1.

Transgene-Containing Plasmid for rAAV Production

The transgene-containing plasmid for rAAV production is engineered to carry at least one heterologous nucleic acid sequence of interest (e.g., an anti-sense RNA molecule, shRNA, miRNA, a ribozyme, or a gene encoding a polypeptide of interest) in which the internal portion of the AAV genome is replaced with a heterologous nucleic acid sequence of interest within an expression cassette. "Expression cassette" as used herein means a nucleic acid sequence capable of directing expression of a particular heterologous nucleic acid sequence in an appropriate host cell (e.g., mammal), which may include a promoter operably linked to the nucleic acid sequence of interest that may be operably linked to termination signals. The expression cassette including the heterologous nucleic acid sequence of interest may be chimeric. The expression cassette may also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression.

In certain embodiments, the transgene-containing plasmid does not comprise an antibiotic resistance gene. In certain embodiments, the transgene-containing plasmid does not comprise an ampicillin resistance gene (e.g., SEQ ID NOS: 71 and 73). While antibiotic resistance genes are commonly used as selection markers for plasmid production, the inclusion of an antibiotic resistance gene (e.g., ampicillin resistance gene) can raise safety concerns. For example, there can be a horizontal gene transfer to patient's bacteria, which would be prevented if the gene is not present in the plasmid. It is particularly important to avoid using antibiotic selection markers involving antibiotics that are in significant clinical use, in order to avoid unnecessary risk of spread of antibiotic resistance traits to environmental microbes (e.g., ampicillin). One should also avoid using antibiotic resistance genes for antibiotics that cause serious hypersensitivity reactions in patients as there could be residual antibiotic in the pharmaceutical composition (e.g., penicillin and other β -lactam antibiotics).

Exemplary transgene-containing plasmids according to the invention is shown in FIGS. 2, 3A, 3B, 15A, and 15B and SEQ ID NOS: 1, 42, 71, and 73, or a plasmid with at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 1, 42, 71, and 73. FIGS. 2, 3A, 3B, 15A, and 15B provides an example of the order of the elements of the transgene-containing plasmids of the invention.

The transgene-containing plasmids according to SEQ ID NOS: 71 and 73 are advantageous because they remove all traces of the ampicillin resistance gene and also include an inactivated gentamycin resistance gene (e.g., the start codon from the open reading frame was removed), which acts as an additional stuffer sequence.

The transgene-containing plasmid is constructed using known techniques to at least provide operatively linked components in the direction of transcription, control elements including a transcriptional initiation region, the DNA of interest and a transcriptional termination region. The control elements are selected to be functional in a mammalian cell. The resulting construct, which contains the operatively linked components, is flanked (5' and 3') with functional AAV inverted terminal repeat (ITR) sequences. Termination signals, such as polyadenylation sites, can also be included in the plasmid.

The ITRs have been shown to be the only cis elements required for packaging allowing for complete gutting of viral genes to create rAAV. Even though the rolling-circle

DNA replication mechanism primarily amplifies (i.e., replicates) the transgene expression cassette DNA sequence flanked by the ITRs due to the presence of the D sequence within the ITRs, the plasmid DNA backbone (e.g., origin of replication, antibiotic resistance gene expression cassette, etc.,) can also be packaged into the vector capsid, albeit at a lower frequency due to the absence of the flanking D sequence domain. AAV is efficient in packaging a genome size similar to or smaller than the wildtype virus genome (~4.7 kbases). One could discourage the packaging of the plasmid backbone by increasing the size of the backbone to such a degree that it is unfavorable for the backbone to be packaged into the capsid. Enlargement of the backbone can be achieved by additional "stuffer" sequences (i.e., filler component), resulting in a plasmid backbone size larger than the wild-type AAV genome. Without wishing to be bound by theory, it is suggested that the presence of an enlarged plasmid backbone can reduce the probability of the rAAV packaging the plasmid backbone into the vector capsid. In some embodiments, the enlarged plasmid backbone is created by use of the stuffer sequence.

In certain embodiments, the stuffer sequence is silent in terms of biological activity, in that it is devoid of at least one of enhancers, promoters, splicing regulators, noncoding RNAs, antisense sequences, and/or coding sequences. In certain embodiments, each of enhancers, promoters, splicing regulators, noncoding RNAs, antisense sequences, and coding sequences are absent.

In certain embodiments, the stuffer sequence comprises an inert intronic DNA sequence found in the human genome. By utilizing a DNA sequence from the human genome, there will be lower probability that the stuffer sequence will elicit an immune response in case the plasmid becomes packaged into the capsid. It is also important that the stuffer sequence does not include an open reading frame.

The stuffer sequence should be large enough that the size of the plasmid backbone is larger than the optimal packaging size of rAAV such that the plasmid backbone is not packaged into the vector capsid. The stuffer sequence can consist of at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 2000, at least 3000, at least 4000, at least 5000, at least 6000, at least 7000, at least 8000, at least 9000 or at least 10000 nucleotides. In certain embodiments, the stuffer sequence comprises a nucleic acid of between 1000 and 5000 nucleotides in length. In certain embodiments, the stuffer sequence comprises a nucleic acid of between 1000 and 2000 nucleotides in length. In certain embodiments, the stuffer sequence comprises a nucleic acid of between 800 and 1500 nucleotides in length. In certain embodiments, the stuffer sequence comprises a nucleic acid of between 800 and 1000 nucleotides in length.

In a preferred embodiment, the stuffer sequence comprises human GAPDH intron 2 (NG007073.2). Without wishing to be bound by theory, the use of human GAPDH intron 2 has lower immunogenicity as it is present in the human genome already and thus should not elicit an immune response if it is by chance packaged. GAPDH intron 2 is ideal as a stuffer sequence as it is a single naturally occurring sequence. There is no need to include any additional nucleotides or to link more than one sequence together, which would result in an unnatural buttressing of DNA sequences.

In certain embodiments, the stuffer sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about

70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 9 or a fragment thereof. In certain embodiments, the stuffer sequence comprises, consists of, or consists essentially of SEQ ID NO: 9 or a fragment thereof.

In certain embodiments, the stuffer sequence comprises an inactivated gentamycin gene. In certain embodiments, the gentamycin gene is modified so that it is not expressed. For example, the start codon could be removed.

In certain embodiments, the stuffer sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 72 or a fragment thereof. In certain embodiments, the stuffer sequence comprises, consists of, or consists essentially of SEQ ID NO: 72 or a non-functional fragment thereof.

The transgene-containing plasmid can be constructed using ITRs from any of the various AAV serotypes. These ITRs base pair to allow for synthesis of the complementary DNA strand. The ITRs remain functional in such plasmids to allow replication and packaging of the rAAV containing the heterologous nucleic acid sequence of interest. Mutations within the terminal repeat sequences of AAV plasmids are well tolerated in generating functional AAV vectors. See e.g., Samulski et al, 1983; Muzychka et al, 1984; and U.S. Pat. No. 9,163,259, which of which as incorporated herein in their entirety for all purposes. Even plasmids with one of the two ITRs deleted, the AAV sequences could be rescued, replicated, and infectious virions be produced, as long as the existing ITR in the construct contains the full AAV ITR sequence.

The nucleic acid sequences of AAV ITR regions are known. The ITR need not have the wild-type nucleic acid sequence, but may be altered, e.g., by the insertion, deletion or substitution of nucleotides. Additionally, the AAV ITR may be derived from any of several AAV serotypes, including without limitation, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, or a chimera thereof. Furthermore, 5' and 3' ITRs which flank a selected nucleic acid sequence in an AAV vector need not necessarily be identical or derived from the same AAV serotype or isolate, so long as they function as intended, i.e., to allow for excision and rescue of the sequence of interest from a host cell genome. Even though SEQ ID NOs: 2 and 43 is used as an example of the 5' ITR sequence of the rAAV described in this document, it is expected that any 5' ITR sequence that carries the terminal resolution site would produce vectors with the same functionality. Likewise, even though SEQ ID NO: 3 is used as an example of the 3' ITR sequence of the rAAV described in this document, it is expected that any 3' ITR sequence that carries the terminal resolution site would produce vectors with the same functionality.

In certain embodiments, the 5' ITR sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 2 or SEQ ID NO 43 or a functional fragment or derivative thereof. In certain embodiments, the 5' ITR comprises, consists of, or consists essentially of SEQ ID NO: 2 or SEQ ID NO 43, or a functional fragment or derivative thereof.

In certain embodiments, the 3' ITR sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 3 or a functional fragment or derivative thereof. In certain embodiments, the 3' ITR comprises, consists of, or consists essentially of SEQ ID NO: 3, or a functional fragment or derivative thereof.

In certain embodiments, the transgene-containing plasmid comprising the stuffer sequence as described above is operably linked to an expression cassette.

In certain embodiments, the expression cassette comprises a promoter. In certain embodiments, the at least one heterologous nucleic acid sequence (e.g., heterologous gene of interest) is operably linked to a pol II promoter (constitutive, cell-specific, or inducible) such that the heterologous nucleic acid sequence is capable of being expressed in the patient's target cells under appropriate or desirable conditions. Numerous examples of constitutive, cell-specific, and inducible promoters are known in the art, and one of skill could readily select a promoter for a specific intended use, e.g., the selection of the muscle-specific skeletal α -actin promoter or the muscle-specific creatine kinase promoter/enhancer for muscle cell-specific expression, the selection of the constitutive CMV promoter for strong levels of continuous or near-continuous expression (e.g., hCMVie (SEQ ID NO: 4)), or the selection of the inducible ecdysone promoter for induced expression. Induced expression allows the skilled artisan to control the amount of protein that is synthesized. In this manner, it is possible to vary the concentration of therapeutic product. Other examples of well-known inducible promoters are: steroid promoters (e.g., estrogen and androgen promoters) and metallothionein promoters. In certain embodiments, the promoter is a pol III promoter. In certain embodiments, the promoter is a U6 promoter. In certain embodiments, the promoter is an H1 promoter. In certain embodiments, the gene expression cassette is without a promoter.

In certain embodiments, the transgene-containing plasmid is multicistronic, i.e., carries more than one gene. Unlike promoters which will create unique mRNA transcripts for each gene that is expressed, multicistronic plasmids simultaneously express two or more separate proteins from the same mRNA. In such cases, the multiple genes are separated by an element that allows for separate translation for each gene (e.g., internal ribosomal entry sites (IRES) or 2A peptides).

Even though SEQ ID NO: 6 is used as an example of an IRES sequence of the rAAV described in this document, it is expected that any 5' ITR sequence that carries the terminal resolution site would produce vectors with the same functionality.

IRES allow for initiation of translation from an internal region of the mRNA by acting as another ribosome recruitment site. In certain embodiments, the IRES sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 6 or a functional fragment or derivative thereof. In certain embodiments, IRES comprises, consists of, or consists essentially of SEQ ID NO: 6 or a functional fragment or derivative thereof.

In certain embodiments, the transgene-containing plasmid encodes a 2A peptide. 2A peptides (see non-limiting examples in Table 1 below) were created to overcome some of the disadvantages of the IRES element. In particular 2A peptides are “self-cleaving” in that these peptides are thought to function by making the ribosome skip the synthesis of a peptide bond at the C-terminus of a 2A element, leading to separation between the end of the 2A sequence and the next peptide downstream. The “cleavage” occurs between the Glycine and Proline residues found on the C-terminus meaning the upstream cistron will have a few additional residues added to the end, while the downstream cistron will start with the Proline. 2A cleavage is universal in eukaryotic cells, and, some scientists report close to 100% cleavage. The choice of specific 2A peptide will ultimately depend on a number of factors such as cell type or experimental conditions, which one of ordinary skill would be understand which one to choose.

capsid proteins). The Rep and/or Cap genes can be derived from AAV serotypes AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and/or AAV11, or a chimera thereof. In certain embodiments, the AAV Rep and/or Cap genes encode genetically engineered AAV and/or chemically modified AAV. See e.g., AAV virions mutated to be less immunogenic such as those recited in U.S. Pat. No. 7,259,151, incorporated herein by reference for all intended purposes. The selection of the AAV serotype can be selected 10 on the tropism of the AAV serotype. Table 2 below provides examples, without limitation, of tropism of the most widely used AAV serotypes. The tropism of AAV can also be modified via pseudotyping (i.e., the mixing of a capsid and genome from ITRs from a different viral serotypes). These 15 serotypes are denoted using a slash, so that AAV2/5 indicates a virus containing the genome carrying ITR of serotype 2 packaged in the capsid from serotype 5. Use of these pseudotyped viruses can improve transduction efficiency, as

TABLE 1

Examples of four common 2A peptides.

Peptide	Amino acid sequence*
T2A:	(GSG) E G R G S L L T C G D V E E N P G P (SEQ ID NO: 74)
P2A:	(GSG) A T N F S L L K Q A G D V E E N P G P (SEQ ID NO: 75)
E2A:	(GSG) Q C T N Y A L L K L A G D V E S N P G P (SEQ ID NO: 76)
F2A:	(GSG) V K Q T L N F D L L K L A G D V E S N P G P (SEQ ID NO: 77)

*(GSG) residues can be added to the 5' end of the peptide to improve cleavage efficiency.

In an embodiment, the plasmid comprises 5' and 3' ITRs from an AAV, wherein the ITRs surround at least one gene. In certain embodiments, a stuffer sequence is located downstream of the 3' ITR. In certain embodiments, the stuffer sequence is upstream of the 5' ITR. ITRs can be from AAV serotypes AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and/or AAV11, or a chimera thereof. In certain embodiments, the ITRs are from AAV serotypes AAV2 and/or AAV5. In certain embodiments, the ITRs can be SEQ ID NO: 2, 3, or 43, or a functional fragment or derivative thereof. In some embodiments, the gene is a reporter gene, such as for example and not limitation, eGFP (e.g., SEQ ID NO: 5) and/or SEAP (e.g., SEQ ID NO: 7). In some embodiments, the stuffer sequence is GAPDH intron 2 or a fragment or variant thereof. In some embodiments, the stuffer sequence is SEQ ID NO: 9 or a fragment thereof. Exemplary gene constructs are shown in FIG. 3 for use in plasmids to generate ssAAV (FIG. 3A) and scAAV (FIG. 3B) rAAV.

Rep-Cap Plasmid

The second plasmid comprises AAV replication (Rep) and capsid (Cap) gene sequences. The AAV Rep-Cap plasmid includes both of the major AAV genes open reading frames (ORFs), Rep gene, and Cap gene. Rep proteins have been shown to possess many functions, including, among others: recognition, binding and nicking of the AAV origin of DNA replication; DNA helicase activity; and modulation of transcription from AAV (or other heterologous) promoters. Cap proteins supply necessary packaging functions and assemble into the viral capsid shell. AAV helper functions are used herein to complement AAV functions in trans that are missing from AAV vectors. Rep and Cap genes are translated to produce multiple distinct proteins (Rep78, Rep68, Rep52, Rep40—required for the AAV life cycle; VP1, VP2, VP3—

well as alter tropism. For example, neurons that are not efficiently transduced by AAV2, one can use AAV2/5, which is distributed more widely in the brain and shown to have improved transduction efficiency. One can also use hybrid capsids derived from multiple different serotypes, which also alter viral tropism. For example, AAV-DJ, which contains a hybrid capsid derived from eight serotypes, displays a higher transduction efficiency *in vitro* than any wild type serotype; *in vivo*, it displays very high infectivity across a broad range of cell types. The mutant AAV-DJ8 displays the properties of AAV-DJ, but with enhanced brain uptake. A number of AAV helper plasmids have been described, such as the commonly used plasmids pAAV/Ad and pIM29+45 which encode both Rep and Cap gene expression products. See, e.g., Samulski et al. (1989) J. Virol. 63:3822-3828; and McCarty et al. (1991) J. Virol. 65:2936-2945 and U.S. Pat. Nos. 5,139,941; 6,001,650; 6,376,237; 7,259,151, each of which are incorporated herein by reference in their entirety for all purposes.

TABLE 2

Tissue Tropism of AAV Serotypes	
Tissue	Optimal Serotype
Heart	AAV1, AAV8, AAV9
Kidney	AAV2
Liver	AAV7, AAV8, AAV9
Nervous System	AAV1, AAV2, AAV4, AAV5, AAV8, AAV9

TABLE 2-continued

Tissue Tropism of AAV Serotypes	
Tissue	Optimal Serotype
Lung	AAV4, AAV5, AAV6, AAV9
Pancreas	AAV8
Photoreceptor Cells	AAV2, AAV5, AAV8
RPE (Retinal Pigment Epithelium)	AAV1, AAV2, AAV4, AAV5, AAV8
Skeletal Muscle	AAV1, AAV6, AAV7, AAV8, AAV9

An exemplary Rep-Cap plasmid according to the invention is shown in FIGS. 4A, 4B, 14A, and 14B; and SEQ ID NOS: 24, 31, 33, 35, 37, 41, 59, and 60, or a plasmid with at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 24, 31, 33, 35, 37, 41, 59, or 60. FIGS. 4A, 4B, 14A, and 14B; provide examples of the order of the elements in the plasmids of AAV Rep-Cap plasmids of the invention.

In certain embodiments, the Rep genes can be derived from AAV serotypes AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and/or AAV11, or a chimera thereof. In certain embodiments, the AAV Rep gene is genetically engineered AAV and/or chemically modified AAV. In certain embodiments, the Rep gene includes genes from AAV serotype 2 (Rep2) and/or Rep5, which includes chimeras (e.g., AAV Rep2/5).

In certain embodiments, the Cap genes can be derived from AAV serotypes AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and/or AAV11, or a chimera thereof. In certain embodiments, the AAV Cap gene is genetically engineered AAV and/or chemically modified AAV. In any of the foregoing embodiments, the Cap gene may be from the same AAV serotype as the Rep gene or a different AAV serotype from the Rep gene. In any of the foregoing embodiments, the plasmid further comprises a Cap gene from any of AAV serotypes 2, 5, 8, and/or 9 (Cap2, Cap5, Cap8, and Cap9, respectively), including chimeric proteins comprising hybrids of Cap proteins from those serotypes.

In certain embodiments, the Rep-Cap plasmid includes, but is not limited to, Rep gene sequence from AAV serotypes 2 and as a chimeric Rep protein combined from more than 1 serotypes, for example Rep2/5, and capsid gene sequence from any AAV capsid serotypes including AAV2, AAV5, AAV8, and/or AAV9.

In certain embodiments, the Rep gene sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 11, 12, 28, or 30, or a functional fragment or derivative thereof. In certain embodiments, Rep gene sequence comprises, consists of, or consists essentially of SEQ ID NOS: 11, 12, 28, or 30, or a functional fragment or derivative thereof.

In certain embodiments, the Cap gene sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 13, 29, 32, or 36, or a functional

fragment or derivative thereof. In certain embodiments, Cap gene sequence comprises, consists of, or consists essentially of SEQ ID NOS: 13, 29, 32, or 36, or a functional fragment or derivative thereof.

5 In certain embodiments, the promoter sequence comprises, consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NO: 34, or a functional fragment or derivative thereof. In certain embodiments, promoter sequence comprises, consists of, or consists essentially of SEQ ID NO: 34, or a functional fragment or derivative thereof.

In an embodiment, the Rep-Cap plasmid further comprises an AAV promoter to control expression of the AAV Rep and Cap proteins described herein. The promoter can be 20 any desired promoter, selected by known considerations, such as the level of expression of a nucleic acid functionally linked to the promoter and the cell type in which the vector is to be used. That is, the promoter can be tissue/cell-specific. Promoters can be prokaryotic, eukaryotic, fungal, 25 nuclear, mitochondrial, viral or plant promoters. Promoters can be exogenous or endogenous to the cell type being transduced by the vector. Promoters can include, for example, bacterial promoters, known strong promoters such as SV40 or the inducible metallothionein promoter, or an 30 AAV promoter, such as an AAV P5 promoter. Additionally, chimeric regulatory promoters for targeted gene expression can be utilized. Examples of these regulatory systems, which are known in the art, include the tetracycline based regulatory system which utilizes the tet transactivator protein (tTA), a chimeric protein containing the VPI 6 activation domain fused to the tet repressor of *Escherichia coli*, the EPTG based regulatory system, the CID based regulatory system, and the Ecdisone based regulatory system. Other 35 promoters include promoters derived from actin genes, immunoglobulin genes, cytomegalovirus (CMV) (e.g., hCMVie (SEQ ID NO: 4), adenovirus, bovine papilloma virus, adenoviral promoters, such as the adenoviral major late promoter, an inducible heat shock promoter, respiratory syncytial virus, Rous sarcomas virus (RSV), etc. The promoter can be the promoter of any of the AAV serotypes and can be the p19 promoter or the p40 promoter. In certain 40 embodiments, the promoter can be an AAV2 P5 promoter or an AAV5 P5 promoter or an AAV P5 promoter. Furthermore, smaller fragments of the P5 promoter that retain promoter activity can readily be determined by standard procedures including, for example, constructing a series of deletions in the P5 promoter, linking the deletion to a reporter gene, and determining whether the reporter gene is expressed, i.e., 45 transcribed and/or translated. Examples of potential promoter can be found in WO2005017101, incorporated by reference herein for all intended purposes. In certain embodiments, the AAV promoter is from AAV serotype 2. Exemplary P5-Rep-Cap plasmids comprising the AAV2 50 promoter P5 are shown in FIGS. 4B and 14B and in SEQ ID NO: 34.

Suitable plasmid backbones for the Rep-Cap plasmid includes but is not limited to, pHLP19, pUC18, pUC19, and pAAV-RC2, see also plasmid backbones described in U.S. 65 Pat. Nos. 6,001,650 and 6,156,303, the entirety of both incorporated herein by reference for all purposes. In certain embodiments, the Rep-Cap plasmid backbone is pUC19.

Ad Helper Plasmid

In an embodiment, the Ad helper plasmid comprises adenovirus genes including, but not limited to, Ad2 and/or Ad5. In an embodiment, the Ad helper plasmid comprises Ad5 genes. The Ad5 gene sequence is used because Ad5 is an efficient helper virus to rAAV. It is known that the full-complement of adenovirus genes is not required for helper function. In fact, it is more desirable to not have the full compliment. For example, adenovirus mutants incapable of DNA replication and late gene synthesis have been shown to be permissive for AAV replication. Ito et al., (1970) J. Gen. Virol. 9: 243; Ishibashi et al., (1971) Virology 45: 317. Thus, the Ad Helper Plasmid is designed to be of minimal size to only carry the required Ad genes required for rAAV production and to serve as a reduced plasmid size construct. It has been shown that adenoviruses defective in the E1 region, or having a deleted E4 region, are unable to support AAV replication. Thus, E1A and/or E4 regions are likely required for AAV replication, either directly or indirectly. Laughlin et al., (1982) J. Virol. 41: 868; Janik et al., (1981) Proc. Natl. Acad. Sci. USA 78: 1925; Carter et al., (1983) Virology 126: 505. Other characterized Ad mutants include: E1B (Laughlin et al. (1982), *supra*; Janik et al. (1981), *supra*; Ostrove et al., (1980) Virology 104: 502); E2A (Handa et al., (1975) J. Gen. Virol. 29: 239; Strauss et al., (1976) J. Virol. 17: 140; Myers et al., (1980) J. Virol. 35: 665; Jay et al., (1981) Proc. Natl. Acad. Sci. USA 78: 2927; Myers et al., (1981) J. Biol. Chem. 256: 567); E2B (Carter, Adeno-Associated Virus Helper Functions, in I CRC Handbook of Parvoviruses (P. Tijssen ed., 1990)); E3 (Carter et al. (1983), *supra*); and E4 (Carter et al. (1983), *supra*; Carter (1995)). Although studies of the accessory functions provided by adenoviruses having mutations in the E1B coding region have produced conflicting results, Samulski et al., (1988) J. Virol. 62: 206-210, recently reported that E1B55k is required for AAV virion production, while E1B19k is not. In addition, International Publication WO 97/17458 and Matshushita et al., (1998) Gene Therapy 5: 938-945, describe accessory proteins encoding various Ad genes. Particularly preferred accessory function plasmids comprise an adenovirus VA RNA coding region, an adenovirus E4 ORF6 coding region, an adenovirus E2A 72 kD coding region, an adenovirus E1A coding region, and an adenovirus E1B region lacking an intact E1B55k coding region. Examples of these plasmids are described in International Publication No. WO 01/83797. Each reference recited in this paragraph are incorporated herein by reference in their entirety for all purposes.

Exemplary Ad helper plasmids according to the invention is shown in FIG. 5 and SEQ ID NOS: 14 and 15, or a plasmid with at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 14 and 15. FIG. 5 provides examples of the order of the elements in the plasmids of Ad helper plasmids of the invention.

In certain embodiments, the Ad helper plasmid can include, without limitation, adenoviral gene sequences for E2a, E4 (orf6), the VA1 RNA gene, and the parvovirus VP capsid gene unit. In certain embodiments, the Ad Helper Plasmid can include VA, E4, and E2A genes. As there is a limitation of how much plasmid can be efficiently transfected to cells for rAAV production, having a reduced size plasmid carrying these Ad genes could help increase the

molar content of all three plasmids used in the transfection, thus increasing the probability of producing higher yield rAAV.

In an embodiment, the Ad helper plasmid comprises E2A, 5 E4 ORFs 1, 2, 3, 4, and 6/7, and VA ("short Ad helper plasmid"). An exemplary short Ad Helper Plasmid is shown in the top panel of FIG. 5. The shorter plasmid described here is to reduce "plasmid load" during the step of transfection so that the overall copy number of plasmids of all 10 three plasmids can be increased to give higher number of plasmid templates for gene expression and replication for rAAV production. The reduced plasmid load is surprisingly useful for larger batches. This may not be a crucial parameter in small research scale production but could be much 15 more critical when scaled up. This exemplary short Ad Helper Plasmid is approximately 12 kb. In another embodiment, the Ad Helper Plasmid comprises E2A, E4 ORFs 1, 2, 3, 4, and 6/7, and VA, as well as genes encoding a protease 20 and a fiber and promoter pVIII ("long Ad helper plasmid"). An exemplary long Ad Helper Plasmid is shown in the bottom panel of FIG. 5. This exemplary long Ad Helper Plasmid is approximately 18 kb.

The differences between the short and long constructs are 25 shown in FIG. 5. The orientations of the three essential gene elements are different. The long version carries additional elements from the adenovirus genome that may have functions that influence rAAV production. The short version contains the minimal gene sequence that is known to be able 30 to support rAAV production.

In certain embodiments, the VA sequence comprises, 35 consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 16 or 48-50, or a functional fragment or derivative thereof. In certain embodiments, VA sequence comprises, 40 consists of, or consists essentially of SEQ ID NOS: 16 or 48-50, or a functional fragment or derivative thereof.

In certain embodiments, the E4 sequence comprises, 45 consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 17, 40, 47, or 55-58, or a functional fragment or derivative thereof. In certain embodiments, E4 sequence comprises, 50 consists of, or consists essentially of SEQ ID NOS: 17, 40, 47, or 55-58, or a functional fragment or derivative thereof.

In certain embodiments, the E2A sequence comprises, 55 consists of, or consists essentially of a nucleic acid having at least about 40%, about 50%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% identity to SEQ ID NOS: 18, 39, 46, or 51, or a functional fragment or derivative thereof. In certain embodiments, E2A sequence comprises, 60 consists of, or consists essentially of SEQ ID NOS: 18, 39, 46, or 51, or a functional fragment or derivative thereof.

Suitable plasmids for the Ad Helper Plasmid include, but 65 are not limited to, pJ241, see also plasmids described in U.S. Pat. Nos. 6,001,650 and 6,156,303, the entirety of both incorporated by reference herein. In certain embodiments, the Ad Helper Plasmid backbone is pUC57.

Additional Genes

In a further embodiment, all three plasmids contain a selection marker. An example of a selection marker includes, but is not limited, to positive selection markers such as drug resistance genes including, but not limited to, G418 (with neor), puromycin (with puror), hygromycin B (with hygr), blasticidin S (with bsrr), mycophenolic acid and 6-thio (guanine) (with gpt) and gancyclovir or 1 (2'-deoxy-2'-fluoro-beta-D-arabinofuranosyl)-5-iodouracil (FIAU) (with HSV-tk), gentamycin, and/or kanamycin (with kanr). In a further embodiment, the drug selection marker on all three plasmids is kanamycin. In certain embodiments, the kanamycin gene comprises or consists of SEQ ID NOs: 19 or 25, or functional fragments or derivatives thereof. In certain embodiments, the gentamycin gene comprises or consists of SEQ ID NOs: 44 or 72, or functional fragments or derivatives thereof.

In an embodiment, one or more of the three plasmids carries one or more reporter genes. Several reporter genes are known in the art and some are commercially available (see, Alam and Cook, *supra*). The reporter gene can be inserted within a plasmid that is particularly suited for an organism and molecular biology manipulations. Promoters of interest can be inserted into cloning sites so that the expression of the reporter gene is under the control of the promoter (see, Rosenthal, N., *Methods Enzymol.* 152: 704-720 (1987); and Shiu, A. and Smith, J. M., *Gene* 67: 295-299 (1988)). Known methods are used to introduce these plasmids into a cell type or whole organism (see, Sambrook et al., *Molecular Biology, A Laboratory Manual*, Cold Spring Harbor Laboratory Press (1989); and Nolan, In: *Molecular Cloning*, Cold Spring Harbor Laboratory Press, (1989)). Examples of reporter genes include, without limitation, β -galactosidase (*LacZ*), firefly luciferase, *Renilla* luciferase, *Gaussia* luciferase, chloramphenicol acetyltransferase (CAT), secreted embryonic alkaline phosphatase (SEAP), cyan fluorescent protein (CFP), green fluorescent protein (GFP), enhanced GFP (eGFP), yellow fluorescent protein (YFP), enhanced YFP (eYFP), blue fluorescent protein (BFP), enhanced BFP (eBFP), red fluorescent protein from the Discosoma coral (*DsRed*), and/or MmGFP (Zemicka-Goetz et al. (1997) *Development* 124: 1133-1137) or others familiar to those of ordinary skill. In another embodiment, one or more of the three plasmids carries a reporter construct comprising both eGFP and SEAP, with an internal ribosome entry site (IRES) located between eGFP and SEAP. In such an embodiment, eGFP, which localizes in the nucleus, can be used for determining vector transduction tropism of the rAAV, while SEAP, which is secreted outside of the cell, can permit quantitative measurement of transduction efficiency, either in culture medium in an in vitro setting, or in the subject's bloodstream in an in vivo setting. *LacZ* can enable color-based selection of desired clones, based on disruption of the *lacZ* gene by a cloned gene.

In an embodiment, each plasmid comprises a unique DNA titer tag. In certain embodiments, the DNA titer tag only appears in the transgene-containing plasmid. In certain embodiments the DNA titer tag appears in all of the plasmid systems. This unique DNA titer tag can be included to enable universal vector genome titering, e.g., via a qPCR (or ddPCR)-based vector genome titering assay, to quantify the amount of vector present. In certain embodiments, the DNA titer tag can be outside the expression cassette but between the 2 ITR's to ensure that it becomes packaged. For example, the DNA titer tag can be upstream of the 3'ITR sequence. As another example, the DNA titer tag can be downstream of the 5'ITR sequence. In certain embodiments,

the DNA titer tag is constructed such that it does not appear endogenously within the subject's genome. For example, the sequence can be compared against the subject's DNA (e.g., via a Blast search or other alignment search tool). The primers used to run the qPCR analysis can also be analyzed to ensure that they do not identify any sequence found in the host cells used to package the virion.

The DNA titer tag can be of a size that allows for efficient qPCR analysis but also takes up the least amount of genome space in the plasmid. In certain embodiments, the DNA titer tag sequence is about 60 nucleotides to about 100 nucleotides in length (e.g. SEQ ID NO: 10) and designed based on a sequence that does not exist in humans or standard laboratory animals. In certain embodiments, the DNA titer tag sequence is about 60 nucleotides to about 80 nucleotides, about 65 nucleotides to about 95 nucleotides, about 70 nucleotides to about 90 nucleotides, or about 75 nucleotides to about 85 nucleotides. In certain embodiments, the DNA titer tag sequence is about 60 nucleotides to about 70 nucleotides, about 65 nucleotides to about 75 nucleotides, about 70 nucleotides to about 80 nucleotides, about 75 nucleotides to about 85 nucleotides, about 80 nucleotides to about 90 nucleotides, about 85 nucleotides to about 95 nucleotides, or about 90 nucleotides to about 100 nucleotides. In certain embodiments, the DNA titer tag sequence is at least about 60 nucleotides, at least about 65 nucleotides, at least about 70 nucleotides, at least about 75 nucleotides, at least about 80 nucleotides, at least about 85 nucleotides, at least about 90 nucleotides, at least about 95 nucleotides, or at least about 100 nucleotides. In certain embodiments, the stretch of DNA titer sequence is 100 nucleotides in length. In certain embodiments, a titer tag of 100 nucleotides can be advantageous in rapid qPCR assays and allow for efficient packaging due to the overall plasmid size and packaging limitations.

Non-limiting examples of nucleic acid sequences that encode DNA titer tags includes SEQ ID NO: 61-70.

Heterologous Nucleic Acid Sequence

Recombinant AAV made by the plasmids of the invention can be administered to one or more cells or tissue of a subject. Thus, the invention embraces the delivery of heterologous nucleic acid sequence that can be useful to modulate the cells or tissue of the subject. For example, rAAV can upregulate or downregulate an activity or product of a cell or tissue.

In certain embodiments, the heterologous nucleic acid sequence can be a heterologous gene of interest encoding one or more peptide, polypeptide, or protein. In certain embodiments, the heterologous nucleic acid sequence can encode a peptide, polypeptide, or protein that binds to a specific target of interest, which can be useful for the treatment or prevention of disease in a subject. Examples of such heterologous nucleic acid sequences and associated peptides, polypeptides, or proteins include, but are not limited to, a gene encoding antibodies, MHC molecules, T-cell receptors, B-cell receptors, aptamers, avimers, receptor-binding ligands, or targeting peptides. Antibodies useful in the present invention can encompass monoclonal antibodies, polyclonal antibodies, antibody fragments (e.g., Fab, Fab', F(ab')², Fv, Fc, etc.), chimeric antibodies, bispecific antibodies, heteroconjugate antibodies, single chain (ScFv), mutants thereof, fusion proteins comprising an antibody portion, humanized antibodies, and any other modified configuration of the immunoglobulin molecule that comprises an antigen recognition site of the required specificity, including glycosylation variants of antibodies, amino acid sequence variants of antibodies, and covalently modified

antibodies. The antibodies may be murine, rat, human, or any other origin (including chimeric or humanized antibodies). An antibody includes an antibody of any class, such as IgG, IgA, or IgM (or sub-class thereof), and the antibody need not be of any particular class. Depending on the antibody amino acid sequence of the constant domain of its heavy chains, immunoglobulins can be assigned to different classes. There are five major classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these may be further divided into subclasses (isotypes), e.g., IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2. The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called alpha, delta, epsilon, gamma, and mu, respectively. The subunit structures and three-dimensional configurations of different classes of immunoglobulins are well known.

In certain embodiments, the heterologous nucleic acid sequence (e.g., heterologous gene of interest) can encode a peptide, polypeptide, or protein that can be useful for the treatment or prevention of disease in a subject. For example, the heterologous nucleic acid sequence can encode a protein X for the treatment of disease Y. Protein X can, for example, substitute for a mutated protein or act to block a mutated protein. Such nucleic acid sequences and associated diseases include, but are not limited to, nucleic acid sequences encoding glucose-6-phosphatase, associated with glycogen storage deficiency type 1A; DNA encoding phosphoenolpyruvate-carboxykinase, associated with Pepck deficiency; DNA encoding galactose-1 phosphate uridyl transferase, associated with galactosemia; DNA encoding phenylalanine hydroxylase, associated with phenylketonuria; DNA encoding branched chain alpha-ketoacid dehydrogenase, associated with Maple syrup urine disease; DNA encoding fumarylacetoacetate hydrolase, associated with tyrosinemia type 1; DNA encoding methylmalonyl-CoA mutase, associated with methylmalonic acidemia; DNA encoding medium chain acyl CoA dehydrogenase, associated with medium chain acetyl CoA deficiency; DNA encoding ornithine transcarbamylase, associated with ornithine transcarbamylase deficiency; DNA encoding arginino-succinic acid synthetase, associated with citrullinemia; DNA encoding low density lipoprotein receptor protein, associated with familial hypercholesterolemia; DNA encoding UDP-glucuronosyltransferase, associated with Crigler-Najjar disease; DNA encoding adenosine deaminase, associated with severe combined immunodeficiency disease; DNA encoding hypoxanthine guanine phosphoribosyl transferase, associated with Gout and Lesch-Nyan syndrome; DNA encoding biotinidase, associated with biotinidase deficiency; DNA encoding alpha-galactosidase-A, associated with Fabry disease; DNA encoding beta-glucocerebrosidase, associated with Gaucher disease; DNA encoding beta-glucuronidase, associated with Sly syndrome; DNA encoding peroxisome membrane protein 70 kDa, associated with Zellweger syndrome; DNA encoding porphobilinogen deaminase, associated with acute intermittent porphyria; DNA encoding alpha-1 antitrypsin for treatment of alpha-1 antitrypsin deficiency (emphysema); DNA encoding C1-esterase for the treatment of hereditary angioedema (HAE); DNA encoding phenylalanine hydroxylase for the treatment of phenylketonuria; DNA encoding acid alpha-glucosidase for the treatment of Pompe disease; DNA encoding ATP7B for the treatment of Wilson's disease; DNA encoding alpha-L-iduronidase for the treatment of mucopolysaccharidose type I (MPSI); DNA encoding iduronate sulfatase for the treatment of mucopolysaccharidose type II (MPSII); DNA encoding heparan sulfamidase for the treatment of

mucopolysaccharidose type IIIA (MPSIIIA); DNA encoding N-acetylglucosaminidase for the treatment of mucopolysaccharidose type IIIB (MPSIIIB); DNA encoding heparan-alpha-glucosaminide N-acetyltransferase for the treatment of mucopolysaccharidose type IIIC (MPSIIIC); DNA encoding N-acetylglucosamine 6-sulfatase for the treatment of mucopolysaccharidose type IIID (MPSIID); DNA encoding galactose-6-sulfate sulfatase for the treatment of mucopolysaccharidose type IVA (MPSIVA); DNA encoding beta-galactosidase for the treatment of mucopolysaccharidose type IVB (MPSIVB); DNA encoding N-acetylgalactosamine-4-sulfatase for the treatment of mucopolysaccharidose type VI (MPSVI); DNA encoding beta-glucuronidase for the treatment of mucopolysaccharidose type VII (MPSVII); DNA encoding hyaluronidase for the treatment of mucopolysaccharidose type IX (MPSIX); DNA encoding erythropoietin for treatment of anemia due to thalassemia or to renal failure; DNA encoding vascular endothelial growth factor, DNA encoding angiopoietin-1, and DNA encoding fibroblast growth factor for the treatment of ischemic diseases; DNA encoding thrombomodulin and tissue factor pathway inhibitor for the treatment of occluded blood vessels as seen in, for example, atherosclerosis, thrombosis, or embolisms; DNA encoding aromatic amino acid decarboxylase (AADC), and DNA encoding tyrosine hydroxylase (TH) for the treatment of Parkinson's disease; DNA encoding the beta adrenergic receptor, DNA encoding anti-sense to, or DNA encoding a mutant form of, phospholamban, DNA encoding the sarco(endo)plasmic reticulum adenosine triphosphatase-2 (SERCA2), and DNA encoding the cardiac adenylyl cyclase for the treatment of congestive heart failure; DNA encoding a tumor suppressor gene such as p53 for the treatment of various cancers; DNA encoding a cytokine such as one of the various interleukins for the treatment of inflammatory and immune disorders and cancers; DNA encoding dystrophin or minidystrophin and DNA encoding utrophin or miniutrophin for the treatment of muscular dystrophies; DNA encoding ABCA4 for the treatment of Stargardt's disease; and, DNA encoding insulin for the treatment of diabetes.

In certain embodiments, the heterologous nucleic acid sequence (e.g., heterologous gene of interest) can encode a peptide, polypeptide, or protein that encodes a blood coagulation protein, which proteins may be delivered to the cells of a subject having a blood disorder (e.g., hemophilia). Examples of such nucleic acids and associated peptides, polypeptides, or proteins include, but are not limited to, DNA encoding Factor IX to a subject for treatment of hemophilia B, Factor VIII to a subject for treatment of hemophilia A, Factor VII for treatment of Factor VII deficiency, Factor X for treatment of Factor X deficiency, Factor XI for treatment of Factor XI deficiency, Factor XIII for treatment of Factor XIII deficiency, and Protein C for treatment of Protein C deficiency.

The invention also includes the expression of engineered artificial DNA binding domain peptide, transcriptional activator or transcriptional repressor and nucleases that can interact with the host cell genome to affect up or down gene expression level for genetic and/or acquired diseases.

The invention also includes the expression of heterologous nucleic acid sequences, including but not limited to, antisense, siRNA, shRNA, miRNA, EGSS, gRNA, sgRNA, ribozymes, or aptamers, which could interact with cellular DNA, RNA and/or proteins that can change the gene expression or activities of proteins for genetic and/or acquired diseases.

The invention also includes the expression of intermediate and/or critical raw material for cellular therapy, including but not limited to rAAV, to be used to infect cells to generate genetically engineered cell therapy materials or drug product.

The invention also includes a heterologous nucleic acid sequence that is a gene editing molecule used for modifying a genomic locus of interest (i.e., target) in a cell. Such modifications include, but are not limited to a disruption, deletion, repair, mutation, addition, alteration, or modification of a gene sequence at a target locus in a gene. Examples of gene-editing molecules include, but are not limited to, endonucleases such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), meganucleases, restriction endonucleases, recombinases, and Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) proteins.

Delivery of rAAV

Recombinant AAV described herein, can be used at a therapeutically useful concentration for the treatment and/or prevention of a disease of interest, by administering to a subject in need thereof, an effective amount of the rAAV made by the plasmids of the invention. Subjects to be treated with rAAV made by the plasmids of the present invention can also be administered with other therapeutic agents or devices with known efficacy for treating or preventing the disease.

Delivery of the rAAV to a subject may be by intramuscular injection or by administration into the bloodstream of the subject. Administration into the bloodstream may be by injection into a vein, an artery, or any other vascular conduit the mutant virions into the bloodstream by way of isolated limb perfusion, a technique well known in the surgical arts, the method essentially enabling the artisan to isolate a limb from the systemic circulation prior to administration of the rAAV. Moreover, for certain conditions, it may be desirable to deliver the mutant virions to the CNS of a subject. By "CNS" is meant all cells and tissue of the brain and spinal cord of a vertebrate. Thus, the term includes, but is not limited to, neuronal cells, glial cells, astrocytes, cerebrospinal fluid (CSF), interstitial spaces, bone, cartilage, intracerebral ventricular, intracranial, cisterna magna injection, intrathecal, intracarotid, intranasal and the like. rAAV or cells transduced in vitro may be delivered directly to the CNS or brain by injection into, e.g., the ventricular region, as well as to the striatum (e.g., the caudate nucleus or putamen of the striatum), spinal cord and neuromuscular junction, or cerebellar lobule, with a needle, catheter or related device, using neurosurgical techniques known in the art, such as by stereotactic injection. See, e.g., Stein et al., J Virol 73:3424-3429, 1999; Davidson et al., PNAS 97:3428-3432, 2000; Davidson et al., Nat. Genet. 3:219-223, 1993; and Alisky and Davidson, Hum. Gene Ther. 11:2315-2329, 2000, each of which are incorporated herein in their entirety for all purposes. For administration to the eye, methods can include, subretinal, intravitreal, trans-scleral, or intracranial.

TABLE 3

Exemplary Sequences for Use in the Plasmids of the Invention	
Sequence ID	Description
SEQ ID NO: 1	Single-strand ITR transgene plasmid
SEQ ID NO: 2	5' ITR
SEQ ID NO: 3	3' ITR
SEQ ID NO: 4	hCMVie
SEQ ID NO: 5	eGFP
5	
SEQ ID NO: 6	IRES
SEQ ID NO: 7	SEAP
SEQ ID NO: 8	SV40 poly A
SEQ ID NO: 9	GAPDH stuffer sequence
SEQ ID NO: 10	DNA titer tag
10	
SEQ ID NO: 11	Rep2/5
SEQ ID NO: 12	Rep2
SEQ ID NO: 13	Cap9
SEQ ID NO: 14	Complete short ad helper plasmid
SEQ ID NO: 15	Complete long Ad helper plasmid
SEQ ID NO: 16	VA gene
SEQ ID NO: 17	E4 gene
SEQ ID NO: 18	E2A
SEQ ID NO: 19	Kanamycin resistance gene (complement)
SEQ ID NO: 20	pUC origin (complement)
SEQ ID NO: 21	P_Amp (complement)
SEQ ID NO: 22	Term_bla (complement)
SEQ ID NO: 23	Rpo_C (complement)
SEQ ID NO: 24	pRep2/5-Cap5 plasmid
SEQ ID NO: 25	Kanamycin resistance gene (complement)
SEQ ID NO: 26	origin of replication; RNaseH cleavage point
SEQ ID NO: 27	lac promoter (complement)
SEQ ID NO: 28	Rep5 BamHI fragment (complement)
SEQ ID NO: 29	Cap5/VP1 fragment (complement)
SEQ ID NO: 30	Start site of Rep2 (complement)
SEQ ID NO: 31	pRep2-Cap2 plasmid
SEQ ID NO: 32	Cap2
SEQ ID NO: 33	pUC19-Kan-Rep2Cap2 plasmid
SEQ ID NO: 34	AAV2/P5 promoter (complement)
SEQ ID NO: 35	pRep2-Cap8 plasmid
30	
SEQ ID NO: 36	Cap8 gene (complement)
SEQ ID NO: 37	pRep2-Cap9 plasmid
SEQ ID NO: 38	lac promoter
SEQ ID NO: 39	E2A gene (complement)
SEQ ID NO: 40	E4 gene (complement)
SEQ ID NO: 41	pUC19-Kan-Rep2Cap9 plasmid
SEQ ID NO: 42	Self-complementary ITR transgene plasmid
SEQ ID NO: 43	Truncated 5' ITR
SEQ ID NO: 44	Gentamycin resistance gene (complement)
SEQ ID NO: 45	pHelper plasmid
SEQ ID NO: 46	E2A complement
SEQ ID NO: 47	E4 (complement)
SEQ ID NO: 48	VA (complement)
SEQ ID NO: 49	VA2 RNA (complement)
SEQ ID NO: 50	VA1 RNA (complement)
SEQ ID NO: 51	E2A-BP
SEQ ID NO: 52	L4 100K
45	
SEQ ID NO: 53	hAdV2, 33-100kD
SEQ ID NO: 54	Incomplete L4 pVII
SEQ ID NO: 55	E4 orf 6/7 (complement)
SEQ ID NO: 56	E4 orf 4 (complement)
SEQ ID NO: 57	E4 orf 3
SEQ ID NO: 58	E4 orf 2 (complement)
SEQ ID NO: 59	pAAV-RC
SEQ ID NO: 60	pUC19-Kan-Rep2Cap8 plasmid
SEQ ID NO: 61	DNA Titer Tag
SEQ ID NO: 62	DNA Titer Tag
SEQ ID NO: 63	DNA Titer Tag
SEQ ID NO: 64	DNA Titer Tag
50	
SEQ ID NO: 65	DNA Titer Tag
SEQ ID NO: 66	DNA Titer Tag
SEQ ID NO: 67	DNA Titer Tag
SEQ ID NO: 68	DNA Titer Tag
SEQ ID NO: 69	DNA Titer Tag
SEQ ID NO: 70	DNA Titer Tag
55	
SEQ ID NO: 71	Single-strand ITS transgene plasmid
SEQ ID NO: 72	Gentamycin resistance gene-inactivated (complement)
SEQ ID NO: 73	Self-complementary ITS transgene plasmid
SEQ ID NO: 74	T2A
SEQ ID NO: 75	P2A
SEQ ID NO: 76	E2A
60	
SEQ ID NO: 77	F2A
65	

TABLE 3-continued

The present disclosure is also described and demonstrated by way of the following examples. However, the use of these and other examples anywhere in the specification is illustrative only and in no way limits the scope and meaning of the disclosure or of any exemplified term. Likewise, the disclosure is not limited to any preferred embodiments described here. Indeed, many modifications and variations of the disclosure may be apparent to those skilled in the art upon reading this specification, and such variations can be made without departing from the disclosure in spirit or in scope. The disclosure is therefore to be limited only by the terms of the appended claims along with the full scope of equivalents to which those claims are entitled.

Example 1: In Vitro Expression of Cap Proteins

This example investigated the in vitro expression of the capsid proteins in AAV293 (Agilent) cells from pUC19-based plasmids as compared to expression levels of the same capsid proteins from control pAAV-RC2 based plasmids carrying Rep2 and Cap2 genes (Agilent) (FIG. 4A).

A first set of four plasmids with varied Rep and Cap genes were created in the pAAV-RC background, starting with Rep2Cap2-pAAV-RC (i.e., pAAV-RC2 as shown in FIG. 4A). The pAAV-RC2 was used to generate Rep2/5Cap5-pAAV-RC, Rep2Cap8-pAAV-RC, and Rep2Cap9-pAAV-RC (FIG. 4A).

A second set of four plasmids were created in the pUC19-Kan background, with the same replication and capsid proteins as the first set. Thus, Rep2Cap2-pUC19-Kan, Rep2/5Cap5-pUC19-Kan, Rep2Cap8-pUC19-Kan, and Rep2Cap9-pUC19-Kan (FIGS. 4A and 14A).

For the experiments, each of the 8 plasmids were separately transfected along with the Ad Helper plasmid, pHHelper (Agilent) (e.g., SEQ ID NO: 45), at a ratio of 1:1.

Expression levels of the Cap protein from the pUC19-Kan-based plasmids were compared to expression levels of the same Cap protein from the pAAV-RC2-based plasmids via Western blotting using a monoclonal B1 antibody (FIG. 6). Positive controls AAV2 reference standard material (RSM) and AAV8 RSM are reference standard materials containing the AAV2 and AAV8 Cap proteins, while the negative control was a cell lysate from HEK293 without any Cap-bearing plasmid. The expression level of capsid protein was AAV5>AAV8>AAV9>AAV2 for both the pUC19-based plasmids and the pAAV-RC2-based plasmids.

FIG. 7 is a Western blot analysis conducted with reduced sample to specifically analyze the amounts of Cap proteins VP1-VP3 more clearly.

Next, the AAV2 P5 promoter was added to the Rep2Cap2 pUC19-Kan, Rep2Cap8 pUC19-Kan, and Rep2Cap9 pUC19-Kan plasmids (e.g., FIGS. 4B and 14B). FIG. 8 shows expression levels of Cap proteins from AAV serotypes 2, 8 and 9 expressed using the P5 promoter tested under the same conditions as described above in comparison with those without the P5 promoter. It was found that the P5 promoter gives higher levels of capsid protein expression. The transgene-containing plasmid and Ad helper plasmid were administered at a ratio of 1:1.

Example 2: Functional Testing of Short and Long Ad Helper Plasmids

The purpose of this example was to test the function of a short Ad helper plasmid and a long Ad helper plasmid versus

the commercial pHHelper. Each plasmid was tested individually to ensure each one is functional before using them in combination.

FIG. 9 shows the positive test results using the short Ad helper plasmid (SEQ ID NO: 14) in the HEK293 host cell system. The short Ad helper plasmid (SEQ ID NO: 14) was tested by co-transfecting the short helper plasmid along with a pTRUF11 transgene-containing plasmid carrying GFP as the transgene between the ITRs and the Agilent RC2 plasmid carrying AAV Rep2 and Cap2 genes. The negative controls consisted of 1) a commercial Ad helper plasmid (pHelper) and the Agilent plasmid RC2 and 2) pHelper and pTRUF11; the positive control consisted of pHelper, pTRUF11, and Agilent RC2 plasmid. After 48 hours, the HEK293 cells were lysed with Triton X-100 and treated with benzonase nuclease to degrade DNA and RNA. The cell lysates, containing AAV particles, were treated with DNase I and serially diluted before undergoing qPCR to determine the viral genome copy number per ml cell lysate. FIG. 9 shows the results of the qPCR assay, with columns 1 and 2 representing the negative controls, column 3 showing the positive control, and column 4 showing the viral genome copy number obtained when the short Ad helper plasmid was used together with 2 other plasmids to produce rAAV.

A similar experiment was performed to test a long Ad helper plasmid (SEQ ID NO: 15) according to the disclosure (FIG. 10). FIG. 10 shows the viral genome copy number per ml cell lysate, as determined by qPCR, of a negative control (column 1), a positive control (column 2), the short Ad helper plasmid (SEQ ID NO: 14)+Rep-Cap bearing plasmid+ITR-GFP bearing plasmid (column 3), and the long Ad Helper Plasmid+Rep-Cap bearing plasmid+ITR-GFP bearing plasmid (column 4). Therefore, the long Ad Helper Plasmid also resulted in the production of AAV.

Example 3: rAAV Virion Production Using the Triple-Plasmid System

The ability of the single strand (ss)- and self-complementary (sc)-ITR-bearing plasmids according to the disclosure to form rAAV virions was tested. In this experiment, the plasmids were co-transfected into HEK293 cells (Agilent). For each transfection, Ad-helper plasmid, Rep-Cap plasmid, and transgene-containing plasmid were used at 1:1:1 molar ratio and 10 ug of total DNA was used per 10 cm plate. The negative control was the commercially available Ad helper plasmid (Agilent) and the commercially available Rep-Cap-bearing plasmid (Agilent), while the positive control used a different ITR-bearing plasmid (ATCC). The top panel of FIG. 11 shows the viral genome copy number per ml cell lysate for the ss-ITR-bearing plasmid (measured by qPCR as above), while the bottom panel shows the viral genome copy number per ml cell lysate for the sc-ITR-bearing plasmid. In both panels, column 1 shows the copy number for the negative control, column 2 represents the positive control, and column 3 represents the plasmid according to the disclosure.

Next, three plasmids according to the disclosure were co-transfected into HEK293 cells (again, 1:1:1 ratio) for qPCR assays. The negative control (column 1 of FIG. 12) was a commercially available Ad helper plasmid and a ITR-bearing plasmid. The positive control (column 2 of FIG. 12) included a commercially available Rep-Cap-bearing plasmid. Columns 3-6 correspond to AAV genome copy numbers from cells transfected with the same commercially available Ad helper plasmid and ITR-bearing plasmid along with a pUC19-based plasmid encoding Rep and Cap pro-

33

teins from AAV serotypes 2, 5, 8, or 9 (noted across the top of the figure). Columns 7-10 of FIG. 12 correspond to AAV genome copy numbers from cells transfected with Ad helper plasmids, pUC19-based Rep-Cap-bearing plasmids, and ITR-bearing plasmids according to the disclosure. Column 11 of FIG. 12 is another positive control corresponding to AAV genome copy number from cells transfected with an Ad helper plasmid and ss-ITR plasmid according to the disclosure and a commercially available Rep-Cap-bearing plasmid.

Example 4: Purification and Production of rAAVs

HEK293 cells were co-transfected with a plasmid system comprising three plasmids according to the disclosure. The cells were chemically lysed, and the cell pellet and medium were collected. The cell lysate was clarified and treated with benzonase. The clarified lysate was run on an appropriate affinity column (e.g., for a plasmid system comprising AAV8 capsid, the affinity column was AVB; for a plasmid system comprising AAV9, the affinity column was AAV9-POROS CaptureSelect). Following a buffer exchange, the rAAV was eluted from the column. The rAAV was then characterized, by way of example and not limitation, by qPCR to determine the viral genome copy number (see FIGS. 9-13, 16). The rAAV can further be evaluated by silver stain to determine purity and identity, by Limulus amebocyte lysate (LAL) assay to measure endotoxin activity and microbial contamination, and by an in vitro transduction assay to determine biological activity. Other characterization assays include alkaline electrophoresis to test the size and integrity of the viral genome, ELISA to examine the capsids, infectious center assays to determine the infectivity of the rAAV particles, and electron microscopy to observe the rAAV particles. Western blotting for specific proteins may also be performed by using appropriate antibodies (see FIGS. 6-8).

Example 5: Use of Tag to Titer Vector Genome

While sequences such as polyA sequences can be used for qPCR quantification, it is not ideal to use such sequences for universal titration. For example, each transgene may use a different polyA sequence (e.g., SV40, bGH polyA, etc.), thereby precluding its use to quantitate vectors across all transgene platforms. Therefore, a separate DNA titer tag outside the transgene cassette (i.e., not transcribed as part of the transgene mRNA transcript) was tested for its ability to universally quantitate any transgene cassette.

A 100 nucleotides DNA titer tag was included upstream of the 3' ITR sequence. This same titer tag can be used in any transgene-containing plasmid for rAAV production to allow

34

for universal vector genome titration via qPCR techniques, which can be used as a single reference standard for any project. The qPCR titration results were compared for the same batch of AAV using either SV40 polyA or the 100 nucleotides DNA titer tag as the target sequence.

Two different viral vectors: rAAV8-ssITR (SEQ ID NO: 1) and rAAV8-scITR (SEQ ID NO: 42) were produced with the transgene-containing plasmid being either single-stranded (SEQ ID NO: 1) or self-complementary (SEQ ID NO: 42) transgene-containing plasmids. Similar qPCR titers were obtained using the two different target sequences, indicating the 100 nucleotides DNA titer tag works equally well as the SV40 polyA, which has been widely used in the field for qPCR-based vector titration (FIG. 13A (rAAV8-ssITR) and 13B (rAAV8-scITR)).

Example 6: Use of Tag to Titer Vector Genome

To further confirm the utility of the DNA titer tag, the same 100 nucleotides DNA titer tag used in Example 5 was included upstream of the 3' ITR sequence in two additional viral vectors: rAAV9-ssITR (SEQ ID NO: 71) and rAAV9-scITR (SEQ ID NO: 73).

While several possible embodiments are disclosed above, embodiments of the present disclosure are not so limited. These exemplary embodiments are not intended to be exhaustive or to unnecessarily limit the scope of the disclosure, but instead were chosen and described in order to explain the principles of the present disclosure so that others skilled in the art may practice the disclosure. Indeed, various modifications of the disclosure in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims. Further, the terminology employed herein is used for the purpose of describing exemplary embodiments only and the terminology is not intended to be limiting since the scope of the various embodiments of the present disclosure will be limited only by the appended claims and equivalents thereof. The scope of the disclosure is therefore indicated by the following claims, rather than the foregoing description and above-discussed embodiments, and all changes that come within the meaning and range of equivalents thereof are intended to be embraced therein.

All patents, applications, publications, test methods, literature, and other materials cited herein are hereby incorporated by reference in their entirety as if physically present in this specification.

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 77

<210> SEQ ID NO 1
<211> LENGTH: 10630
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6316)..(6415)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:

```

-continued

<221> NAME/KEY: misc_feature
 <222> LOCATION: (6316)..(6415)
 <223> OTHER INFORMATION: This region may encompass 60-100 nucleotides
 <400> SEQUENCE: 1

tagggaaata	ggccagggtt	tcaccgtAAC	acgccacATC	ttgcgaatAT	atgtgttagAA	60
actgcccggaa	atcgtcgTGT	gcactcatGG	aaaacggTGT	aacaaggGtg	aacactatCC	120
catatcacca	gctcaccgTC	tttattGcc	atacggAACT	ccggatgAGC	attcatcagg	180
cgggcaagaa	tgtgaataAA	ggccggataA	aacttgtGCT	tatTTTCTT	tacggTCTT	240
aaaaaggccg	taatatccAG	ctgaacggTC	tggTTatAGG	tacattgAGC	aactgactGA	300
aatgcctaa	aatgttCTT	acgatGCCAT	tgggataATA	caacggTGGT	atatccAGTG	360
atTTTTCT	ccatTTTT	ttcctcTTT	agaaaaACTC	atcgagcATC	aatgaaACT	420
gcaatttatt	catatcAGGA	ttatcaataAC	catatTTTG	aaaaaggccgT	ttctgtAATG	480
aaggagaaaa	ctcaccgagg	cagttccata	ggatggCAAG	atcctggTAT	cggTctgcGA	540
ttccgactcg	tccaacatCA	atacaaccta	ttaatttccc	ctcgtaaaaa	ataaggTTAT	600
caagtgagaa	atcaccatGA	gtgacgactG	aatccggTGA	gaatggCAAA	agtttatGCA	660
tttcttcca	gacttgttCA	acaggccAGC	cattacgTC	gtcatcaAAA	tcactcgcAT	720
caaccaaacc	gttattcatt	cgtgattgcG	cctgagcgAG	gcgaaatacG	cgatcgCTGT	780
taaaaggaca	attacaAAACA	ggaatcgAGT	gcaaccggCG	caggaacACT	gccagcGCAT	840
caacaatatt	ttcacctgAA	tcaggatatt	cttctaataC	ctggAACGCT	gttttccGG	900
ggatcgcagt	ggtgagtaAC	catgcATCAT	caggagtACG	gataAAATGC	ttgatggTCG	960
gaagtggcat	aaattccgTC	agccagTTA	gtctgaccAT	ctcatctGTA	acatcatTGG	1020
caacgcTacc	tttgcCATGT	ttcagaaACA	actctggCGC	atcgggCTC	ccatacaAGC	1080
gatagattgt	cgcacCTGAT	tgcccGACAT	tatcgCGAGC	ccatttataAC	ccatataAAAT	1140
cagcatccat	gttggAAATT	aatcgccgCC	tcgacgttTC	ccgttGAATA	tggctcATT	1200
tttttccT	ctttaccaAT	gtttaatcAG	tgaggcacCT	atctcagcGA	tctgtctATT	1260
tcgttcatcc	atagttgcCT	gactccccGT	cgtgtagata	actacgatac	gggaggGCTT	1320
accatctggc	cccagcgCTG	cgatgataCC	gcgagaACCA	cgctCACCGG	ctccggATT	1380
atcagcaata	aaccagccAG	ccggaaggGC	cgagcgcAGA	agtggcCTG	caactttATC	1440
cgcctccatc	cagtctatta	attgttgcCG	ggaagCTAGA	gtaagttagTT	cgccagTTAA	1500
tagttgcgc	aacgttGTTG	ccatcgCTAC	aggcatcgTG	gtgtcacGCT	cgtcgTTGG	1560
tatggcttca	ttcagctCCG	gttcccAACG	atcaaggcGA	gttacatGAT	cccccatGTT	1620
gtgcacgttg	tcagaagtaA	gttggccGCA	gtgttatCAC	tcatggTTAT	ggcagcACTG	1680
cataattctc	ttactgtCAT	gccatccGTA	agatgCTTT	ctgtgactGG	tgagtactCA	1740
accaagtcat	tctgagaATA	gtgtatgcGG	cgaccgAGT	gtcttgcCC	ggcgtcaATA	1800
cgggataata	ccgcgcCACA	tagcagaACT	ttaaaAGTGC	tcatcattGG	aaaacgttCT	1860
tcggggcgaa	aactctcaAG	gatTTACCG	ctgttgAGAT	ccagttcgAT	gtAAcccACT	1920
cgtgcaccca	actgatCTTC	agcatCTTT	actttcacCA	gcgtttCTGG	gtgagcaAAA	1980
acaggaaggc	aaaatGCCGc	aaaaaaggGA	ataaggGCgA	cacggAAATG	ttgaataACTC	2040
atattctcc	ttttcaata	ttattgaAGC	atttatCAGG	gttattgtCT	catgagcGGA	2100
tacatatttG	aatgtattta	aaaaataAA	caaataGGGG	tcagtgttAC	aaccaatTA	2160
ccaaattctga	acattatcgc	gagcccATT	atacctGAAT	atggctcATA	acacccCTTG	2220

-continued

tttgctggc ggcagtagcg cggtggtccc acctgacccc atgccgaact cagaagtcaa	2280
acgcctgac gccgatggta gtgtgggac tccccatgcg agagtagggaa actgccaggc	2340
atcaaaaacgaaaggct cagtcgaaag actgggcctt tcgccccggc taattgaggg	2400
gtgtcgccct tattcgactc ggggctcgag ctgcgcgcctc gtcgtgtcac tgaggccgccc	2460
cgggcaaagc cggggcgctcg ggccgaccttt ggtcgccccgg cctcagttag cgagcgagcg	2520
cgcagagagg gagtgcccaa ctccatcaact aggggttccct ttaattaaac gcgtttacat	2580
aacttacggt aaatggcccg cctggctgac cgcccaacgaa ccccccggcca ttgacgtcaa	2640
taatgacgta tgttccata gtaacgcca tagggacttt ccattgacgt caatgggtgg	2700
actatattacg gtaaactgccc cacttggcag tacatcaagt gtatcatatg ccaagtacgc	2760
cccttattgtc cgtcaatgac ggttaatggc cccgcctggca ttatgcggcag tacatgaccc	2820
tatggactt tcctacttgg cagtcacatct acgttattatg catcgctatt accatgggtga	2880
tgccgttttg cagtcacatc aatgggcgtg gatagcggtt tgactcacgg ggattccaa	2940
gtctccaccc cattgacgtc aatgggagtt tgtttggca caaaaatccaa cgggacttcc	3000
aaaaatgtcg taacaactcc gccccattga cgccaaatgggg cggtaggcgt gtacgggtgg	3060
aggctatatc aggccgcgcg aactgaaaaaa ccagaaagt aactggtaag ttttagtcttt	3120
ttgtctttta ttcagggtcc cggatccggt ggtggcata atcaaagaac tgctcttcag	3180
tggatgttgc cttaacttctt aggcctgtac ggaagtgtta ctctgtctt aaaagctctt	3240
gcagggaatt cggcaccatcg gtgagcaagg gcggaggagct gttcacccggg gtggcggcca	3300
tcctggtcga gctggacggc gacgtaaacg gccacaagtt cagcgtgtcc ggccggcgg	3360
agggcgtgc cacctacggc aagctgaccc tgaagttcat ctgcaccacc ggcaagctgc	3420
ccgtgcctg gcccacccctc gtgaccaccc tgacctaagg cgtgcagtgc ttccggcgt	3480
accccgacca catgaagcagc cacgacttctc tcaagtcgcg catgccccaa ggctacgtcc	3540
aggagcgcac catcttcttc aaggacgcacg gcaactacaa gacccggcggc gaggtgaagt	3600
tcgagggcga caccctgggt aacccgcateg agctgaaggcatcgacttc aaggaggacg	3660
gcaacatcct ggggcacaag ctggagttaca actacaacag ccacaacgtc tatatcatgg	3720
ccgacaagca aagaacggc atcaagggtga acttcaagat cccgcacaac atcgaggacg	3780
gcagcgtcga gtcgeccac cactaccago agaacacccc catcgccgac ggccccgtgc	3840
tgtgtcccgca caaccactac ctgagcaccc agtccgcctt gagcaagac cccaaacgaga	3900
agcgcgtatca catggccttg ctggagttcg tgaccggcgc cgggatcaact ctggcatgg	3960
acgagctgtatca caagtaatag actagtgcctt ctctccctcc cccccccctta acgttactgg	4020
ccgaaggccgc ttggataaag gcccgtgtgc gtttgttat atgtttttt ccaccatatt	4080
ccgtctttt ggcaatgtga gggccggaa acctggccctt gtcttcttgc cgagcatcc	4140
taggggtctt tccccctctcg ccaaaggaaat gcaaggctcg ttgaatgtcg tgaaggaagc	4200
agtccctctg gaagcttctt gaagacaaac aacgtctgtat cgcacccttt gcaggcagcg	4260
gaaccccccaca cctggcgaca ggtgcctcg cggccaaaggcc acacgtgtat aagatacacc	4320
tgcaaaggcg gcacaacccc agtgcacgt tggatgttgg atagttgtgg aaagagtc	4380
atggctacc tcaagcgtat tcaacaagggt gctgaaggat gcccagaagg tacccatttgc	4440
tatggatctt gatctggggc ctcggcgtac atgttttaca tgtgttttagt cgaggtaaa	4500
aaacgtcttag gccccccgaa ccacggggac gtggtttcc tttgaaaaac acgtatgataa	4560

US 12,383,587 B2

39

40

-continued

taccgggtgcc accatgctgc tgctgctgtc gctgctgggc ctgaggctac agctctccct	4620
ggccatcatc ccagttgagg aggagaaccc ggacttctgg aaccggcgagg cagccggaggc	4680
cctgggtgcc gccaagaagc tgcagcctgc acagacagcc gccaagaacc tcatacatctt	4740
cctgggcat gggatggggg tgtctacggt gacagctgcc aggatctaa aagggcagaa	4800
gaaggacaaa ctggggcctg agatacccct ggccatggac cgcttccat atgtggctct	4860
gtccaagaca tacaatgtag acaaacatgt gccagacagt ggagccacag ccacggccta	4920
cctgtgeggg gtcaaggga acttccagac cattggcttg agtgcagccg cccgtttaa	4980
ccagtgcAAC acgacacgcg gcaacgaggt catctccgtg atgaatcggg ccaagaaagc	5040
agggaagtca gtgggagtgg taaccaccac acgagtgca caccgcctcgc cagccggcac	5100
ctacgcccc accggtaacc gcaactggta ctggacgcg gacgtgcctg cctcgcccg	5160
ccaggagggg tgccaggaca tcgctacgca gctcatctcc aacatggaca ttgacgtat	5220
ccttagtgga ggccgaaagt acatgtttcg catggaaacc ccagaccctg agtacccaga	5280
tgactacagc caaggtggga ccaggctggc cgggagaat ctggtgagg aatggctggc	5340
gaagcgcac ggtgcccgggt atgtgtggaa ccgcactgag ctcatgcagg cttccctgg	5400
cccgctgtg accccatctca tgggtcttctt tgagcctggc gacatgaaat acgagatcca	5460
cccgagactcc acactggacc cctccctgat ggagatgaca gaggctgccc tgccctgct	5520
gagcaggaac ccccgccgct tccctctt cgtggagggt ggtcgcatcg accatggtc	5580
tcatgaaagc agggcttacc gggactgac tgagacgatc atgttcgacg acgccatga	5640
gagggcggc cagctcacca gcgaggagga cacgctgacg ctgcgtactg ccgaccactc	5700
ccacgtcttc tccttcggag gctacccct gcgagggagc tccatctcg ggctggcccc	5760
tggcaaggcc cgggacagga aggctacac ggtcctccta tacggaaacg gtccaggcta	5820
tgtgctcaag gacggcgcacc ggccggatgt taccgagacg gagagcggga gccccgagta	5880
tcggcagcag tcagcagtgc ccctggacga agagaccac gcagggcgagg acgtggcggt	5940
gttcgcgcgc ggcccgcagg cgcacctgggt tacggcgtg caggagcaga cttcatagc	6000
gcacgtcatg gccttcgcgc cctgcctggc gcctacacc gcctgcgacc tggcgcccc	6060
cgccggcacc accgacgccc cgcacccggg ttactctaga gtcggggcgg ccggccgctt	6120
cgagcagaca tgagtcgaca gatctttaa aaaacctccc acacaattgt tgggttaac	6180
ttgtttattt cagttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat	6240
aaagcatttt ttcactgca ttctagttgt ggtttgtcca aactcatcaa tgcgtatcttat	6300
catgtctgtt taaacnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	6360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnattta	6420
aataggaacc cctagtgtat gagttggcca ctccctctat ggcgcgtcgc tcgctactg	6480
aggccggccg accaaaaggc gcccacgcg cgggcttgc cggggggcc tcagtgagcg	6540
aggcggcgccg cagagagat acatcgatgt gagttcgccg gtggctgggg ggccctggc	6600
tgcgaccgc cccgaaccgc gtctacgago cttgcgggtt ccgggtcttt gcagtcgtat	6660
ggggggcaggg tagctgttcc ccgcaaggag agctcaaggt cagcgctcgg acctggcgga	6720
gccccgcacc caggctgtgg cgccctgtgc agctccgccc ttgcggcgcc atctgcccgg	6780
agcccttc ccctagtccc cagaaacagg aggtccctac tcccgccgaa gatcccgacc	6840
cggaccccta ggtggggac gctttcttcc ctttcgcgc ctgcggggc acgtgtcgca	6900
gaggagcccc tcccccacgg cctccggcac cgccaggcccc gggatgctag tgccgcagg	6960

-continued

gtgcatccct gtccggatgc tgccgcgtcg gtagagcggc cgccatgtt caaccggaa 7020
 gggaaatgaat gggcagccgt taggaaagcc tgccgggtgac taaccctgcg ctcctgcctc 7080
 gatgggtgga gtcgcgtgtg gcggggaaagt caggtggagc gaggctagct ggcccattt 7140
 ctcctccggg tgatgtttt cctagattat tctctggtaa atcaaagaag tgggttatg 7200
 gaggtcctct tgggtccccct ccccgagag gtgtggtgcc tgggtcatgg tgccaagccg 7260
 ggagaagctg agtcatgggt agttggaaaa ggacatttc accgc当地 7320
 gtgggtggccc ttccctgcag cgccggctca ctcacggcc ccgc当地 cctgc当地 7380
 tagcgttgac cc当地ccaa aggccaggct gtaaatgtca cc当地gggat tgggtgtctg 7440
 ggccgc当地gg ggaacctgccc ttccgtcttcc cggaaaccag atctccacc 7500
 gcaccctggt ctgaggttaa atatacgctc tgacccccc ttagctgggg gc当地ggctg 7560
 gggctctctc cc当地cccttc tccccacaca catgcactta cctgtgtcc cactctgat 7620
 ttctggaaaa gagctaggaa ggacaggca cttggcaaat caaagccctg ggactagggg 7680
 gttaaaatac agctccccc cttccaccc gcccaggct ctgtccctt tggtaggg 7740
 acttagagaa ggggtggct tgccctgtcc agttaatttc tgacccccc tc当地ccctt 7800
 ttagttttagt gatgtctgat gtacaaggct tttccctta aagggtgcag ctgagctagg 7860
 cagcagcaag catccctggg gtggcatgtt ggggtggtaa ataccatgtt caaagctgt 7920
 gccc当地ctg tgggtggcag tgcccccacat ggc当地cttcc cttggatga 7980
 ctgggggtgt tggcagcccc tggagccctc agttgc当地cc atgc当地taa ccaggccagg 8040
 ctggcaggaa agctcaaggaa agataaaattt cttcccttgg tggcttaggg ctgctcacat attctggagg 8100
 gatgtctcat tc当地ccctt aatggggagg tggcttaggg ctgctcacat attctggagg 8160
 agccctccct cctcatgcct tcttgcctt tgc当地cccttgcatgcaaaa gagtgc当地 8220
 agggc当地ac acaaaattt ctaatgc当地 aataaatact gataacatct tatagttgt 8280
 attataattt gtattatcgat tgacatgtat aattttgata tcaaaaactg attttccctt 8340
 tattatccctt gagatttt ttcttaatcc tctttaacaa actagaataa ttgtatatac 8400
 aaaaaatcat aaataataga tgaatagtt aattatagg tttcatcaat cgaaaaagca 8460
 acgtatctt tttaaagtgc gttgtttt tctcatttt aaggtaaat aattctata 8520
 tatcaagcaa agtgc当地ggc gccc当地aaat attctgacaa atgc当地ttcc cctaaactcc 8580
 ccccaaaaaa aaaccggccg aagcgggttt ttacgttatt tgccgattaa cgattactcg 8640
 ttatcagaac cggccagggg gccc当地gtt aagactggcc gtc当地tttac aacacagaaa 8700
 gagttttagt aaacgc当地aa aggccatccg tcaggccct tctgcttagt ttgtgc当地 8760
 gc当地ccctt actctcgccct tccgttccct cgctcaactga ctgc当地gc当地 tgggtcggtt 8820
 ggctgc当地gg agcggatca gctcaactaa aggccgtat acggttatcc acagaatcag 8880
 gggataacgc agggaaagaac atgtgagca aaggccagca aaaggccagg aaccgtaaaa 8940
 aggccgctt gctggcttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc 9000
 gagcgtcaag tcaagagggtgg cgaaaccgc当地 caggactata aagataaccag ggc当地ccccc 9060
 ctggaagctc ctc当地gc当地 tctcctgtcc cgaccctgc当地 gcttaccggc tacctgtcc 9120
 ccttctccctt ttc当地ggaaagc gtc当地ggctt ctcatagctc acgctgttagg tatctcagtt 9180
 cgggttaggt cgttgc当地ccctt aagctggctt gtc当地gc当地 accccccgtt cagccgc当地 9240
 gctgc当地ccctt atccggtaac ttc当地gttcc agtccaaaccg ggtaaagacac gacttgc当地 9300

-continued

cactggcagc agccactggt aacaggatta gcagagcgag gtagatgtggc ggtgtacag	9360
agttcttcaa gtgggtggct aactacggct acactagaag aacagtattt ggtatctgcg	9420
ctctgctgaa gccagttacc ttccggaaaaa gagttggtag ctcttgatcc ggcaaacaaa	9480
ccaccgctgg tagccgggtt tttttgttt gcaagcagca gattacgcgc agaaaaaaaaag	9540
gatctcaaga agatcccttg atcttttcta cgggggtcga cgctcgtgg aacgacgcgc	9600
gcttaactca cgtaaggaa ttgggtcat gagcttgcgc cgtcccgatca agtcagcgt	9660
atgctctgct taggtggcgg tacttgggc gatatcaaag tgcatcactt cttccgtat	9720
gcccaacttt gtatagagag ccactgcggg atcgtcaccc taatctgcct gcacgttagat	9780
cacataagca ccaagcgcgt tggcctcatg cttgaggaga ttgatgagcg cggggcaat	9840
gcctgcctc cgggtgcgc cggagactgc gagatcatag atatagatct cactacgcgg	9900
ctgctcaaacc ttgggcagaa cgtaagccgc gagagcgcaca acaaccgcctt cttggtcgaa	9960
ggcagcaagc gcgatgaatg tcttactacg gagcaagtcc cggaggtaat cggagtccgg	10020
ctgatgttgg gagtaggtgg ctacgtcacc gaaactcagca cggaaaagat caagagcgc	10080
ccgcatggat ttgacttggg cagggccgag cctacatgtc cgaatgtgc ccataacttga	10140
gccacctaacc ttgttttag ggcgactgcctt ctgctgcgtt acatcggttgc tgctccataa	10200
catcaaaccat cgacccacgg cgtaacgcgc ttgctgcttg gatgcccgg gcatagactg	10260
tacaaaaaaaaa cagtcatataac aagccatgaa aaccgcact ggcgcgttac caccgctgcg	10320
ttcggtaaag gttctggacc agttgcgtga ggcgcattttt ttttctctt cggcggttac	10380
ccccccgcctt gccactcatc gcaactgtt tgtaattcat taaggcattt gcccgcattgg	10440
aagccatcac agacggcatg atgaacactgatc atcgccagcg gcatcagcac cttgtgcct	10500
tgcgctataat atttgcccat agtggaaaacgg ggggcgaaga agttgtccat attggccacg	10560
tttaaatcaa aactggtgaa actcacccag ggattggcgc tgacgaaaaaaaaa catattctca	10620
ataaaaccctt	10630

<210> SEQ ID NO 2
<211> LENGTH: 130
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 2
ctgcgcgcgtc gctcgctcac tgaggccgc cgggcaaaagc cggggcggtcg ggcgacccctt 60
ggtcgcggccgg cctcagttag cgagcgagcg cgcagagagg gagttggccaa ctccatcact 120
agggggttccct 130

<210> SEQ ID NO 3
<211> LENGTH: 133
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 3
aggaacccctt agtggatggag ttggccactc cctctctgcg cgctcgatcg ctcactgagg 60
ccgggcgacc aaaggctgccc cgcacggccgg gctttggccgg ggccgcctca gtgagcgac 120
gagcgccgcag aga 133

-continued

<210> SEQ ID NO 4
<211> LENGTH: 503
<212> TYPE: DNA
<213> ORGANISM: Human cytomegalovirus

<400> SEQUENCE: 4

```
acgcgtttac ataacttacg gtaaatggcc cgcctggctg accgcccac gaccggcc 60
cattgacgtc aataatgacg tatgtccca tagtaacgcg aataggact ttccattgac 120
gtcaatgggt ggactattt cggtaaactg cccacttggc agtacatcaa gtgtatcata 180
tgccaagtac gccccattt gacgtcaatg acggtaatg gcccgectgg cattatgccc 240
agtacatgac cttatggac tttctactt ggcagtagat ctacgtat ttcgtatcgcta 300
ttaccatggt gatgcgggtt tggcagtaca tcaatggcg tggatagccg tttgactcac 360
ggggatttcc aagtctccac cccattgacg tcaatggag tttgtttgg cacaaaatc 420
aacgggactt tccaaaatgt cgtaacaact ccggccattt gacgcaatg ggcggtaggc 480
gtgtacggtg ggaggtctat ata 503
```

<210> SEQ ID NO 5
<211> LENGTH: 723
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 5

```
atggtgagca agggcgagga gctgttcacc ggggtgggtgc ccattctggt cgagctggac 60
ggcgacgtaa acggccacaa gttcagcggtc tccggcgagg gcgaggccga tgccacctac 120
ggcaagctga ccctgaagtt catctgcacc accggcaagc tgccctgccc ctggccacc 180
ctcgtgacca ccctgaccta cggcggtcag tgcttcagcc gctaccggc ccacatgaag 240
cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatttc 300
ttcaaggacg acggcaacta caagacccgc gccgagggtga agttcgaggg cgacaccctg 360
gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420
aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac 480
ggcatcaagg tgaacttcaa gatccgccc acatcgagg acggcagcgt gcagctcgcc 540
gaccactacc agcagaacac ccccatcgcc gacggccccc tgctgtgcc cgacaaccac 600
tacctgagca cccagtcgc cctgagcaaa gaccccaacg agaagcgca tcacatggtc 660
ctgctggagt tcgtgaccgc cgccggatc actctcgcc tggacgagct gtacaagtaa 720
tag 723
```

<210> SEQ ID NO 6
<211> LENGTH: 575
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 6

```
gcccccttc cttttttttt cctaacgtta ctggccgaag ccgcttggaa taaggccgt 60
gtgcgttgtt cttatgtta tttccacca tattgcccgc ttttggcaat gtgagggccc 120
```

-continued

ggaaacctgg ccctgtttc ttgacgagca ttccctagggg tcttccctt ctgcacaaag	180
gaatgcaagg tctgttgaat gtcgtgaagg aagcagtcc tctggaaagct tcttgaagac	240
aaacaacgtc tgtagegacc ctttgcaggg ageggAACCC cccacctggc gacaggtgcc	300
tctgcggcca aaagccacgt gtataagata cacctgaa ggcggcacaa ccccagtggc	360
acgttgttagt ttggatagtt gtggaaagag tcaaattggct cacctcaagc gtattcaaca	420
aggggctgaa ggatgeccag aaggtaaaaa attgtatggg atctgtatcg gggcctcggt	480
gcacatgttt tacatgtgtt tagtgcaggt taaaaaacgt ctagggcccc cgaaccacgg	540
ggacgtggtt ttccttgaa aaacacgtat ataat	575

<210> SEQ ID NO 7

<211> LENGTH: 1560

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 7

atgtgtgtgc tgctgtgtct gctgggcctg aggctacagc tctccctggg catcatccca	60
gtttagggagg agaaccggga ctcttggaaac cgcgaggccag ccggaggccct gggtgccggcc	120
aagaagctgc agcctgcaca gacagccggc aagaacctca tcatcttctt gggcgatggg	180
atgggggtgt ctacggtgac agctgccagg atcctaaaag ggcagaagaa ggacaaactg	240
gggcctgaga taccctggc catggaccgc ttcccatatg tggctctgtc caagacatac	300
aatgttagaca aacatgtgcc agacagtggc gccacagecca cggcctactt gtgcggggtc	360
aaggggcaact tccagaccat tggcttgagt gcagccggcc gctttaacca gtgcacacag	420
acacgcggca acgagggtcat ctccgtgtat aatcgggccca agaaaggcagg gaagtcagt	480
ggagtggtaa ccaccacacg agtgcagcac gcctcgccag cggcgcaccta cgccacacg	540
gtgaaccgca actggactc ggacgcccgcac gtgcctgcct cggccgcaca ggaggggtgc	600
caggacatcg ctacgagct catctccaaat atggacattt acgtgtatctt aggtggaggc	660
cgaaagtaca tggatcgat gggaaacccca gaccctgagt acccagatga ctacagccaa	720
ggtagggacca ggctggacgg gaagaatctg gtgcaggaat ggctggcgaa ggcgcagggt	780
gcccggatgt tggatcgacc cactgagctc atgcaggctt ccctggaccc gtctgtgacc	840
catctcatgg gtctcttgc gcctggagac atgaaatacg agatccaccc agactccaca	900
ctggacccct ccctgtatgg gatgacagag gctgcctgc gcctgtgag caggaacccc	960
cggcgttct tcctttcgat ggagggttgtt cgcacatcgacc atggatcatca tgaaagcagg	1020
gcttaccggg cactgactga gacgatcatg ttgcacgacg ccattgagag ggcggccag	1080
ctcaccagcg aggaggacac gctgagccctc gtcactgccc accactccca cgtcttctcc	1140
ttcggaggct accccctgcg agggagctcc atcttcgggc tggccctgg caaggcccgg	1200
gacaggaagg cctacacggc cctccatac ggaaacggcc caggctatgt gctcaaggac	1260
ggcgccggc cggatgttac cgagagccag agcggggagcc cggagtatcg gcagcagtca	1320
gcagtgcccc tggacgaaag aacccacgca ggcggaggacg tggcggttgtt cgcgcggc	1380
ccgcaggccgc acctgggtca cggcgatgcag gagcagaccc tcatagcgca cgtcatggcc	1440
ttcggccgcct gcctggagcc ctacaccgc tgcgacctgg cggcccccgc cggcaccacc	1500
gacgcccgc acccggttta ctcttagagtc gggcgccgcg gccgcttcga gcagacatga	1560

-continued

<210> SEQ ID NO 8
<211> LENGTH: 168
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 8

agatctttta	aaaaacctcc	cacacaattt	ttgttgttta	cttggttatt	gcaggttata	60
atgggttacaa	ataaaagcaat	agcatcacaa	atttcacaaa	taaagcattt	ttttactgca	120
attcttagttg	tgggttgtcc	aaactcatca	atgtatctta	tcatgtct		168

<210> SEQ ID NO 9

<211> LENGTH: 1632
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 9

gtgagttcg	gggtggctgg	ggggccctgg	gctgcgaccg	cccccaacc	gcgtctacga	60
gccttgcggg	ctccgggtct	ttgcagtcgt	atggggcag	ggtagctgtt	cccccaagg	120
agagctcaag	gtcagcgctc	ggacotggcg	gagccccgca	cccagggtgt	ggcgcctgt	180
gcagctccgc	ccttgcggcg	ccatctgccc	ggagccctct	tcccttagtc	cccagaaaca	240
ggaggteccc	actccggccc	gagatcccga	cccgagcccc	tagtggggg	acgctttctt	300
tccttgcgc	ctctgcgggg	tcacgtgtcg	cagaggagcc	cctcccccac	ggcctccggc	360
acccgcaggcc	ccgggatgt	agtgcgcago	gggtgcaccc	ctgtccggat	gctgegcctg	420
cggttagagcg	gcccgcattgt	tgcaaccggg	aaggaaatga	atgggcagcc	gttagaaag	480
cctgcccgtg	actaaccctg	cgctctgc	tcgatgggtg	gagtcgcgtg	tggcgggaa	540
gtcagggtgga	gcgaggctag	ctggcccgat	tttcctccg	ggtgatgc	ttcttagatt	600
attctctgt	aatcaaaga	agtgggtta	tggaggtct	cttgcgtccc	ctcccccgag	660
agggtgtgt	gtctggcat	ggtccaago	cgggagaagc	ttagtcattgg	gttagttggaa	720
aaggacattt	ccaccgcaaa	atggccctc	tggtggtggc	cccttctgc	agcgcggct	780
cacctcacgg	ccccgcctt	ccctgccag	cctagcggtt	acccgacccc	aaaggccagg	840
ctgtaaatgt	caccggagg	attgggtgtc	tgggcgcctc	ggggAACCTG	cccttctccc	900
cattccgtct	tccggaaacc	agatctccca	ccgcaccctg	gtctgagggtt	aaataatagct	960
gctgacctt	ctgtagctgg	gggcctggc	tggggctctc	tcccatccct	tctccccaca	1020
cacatgcact	tacctgtgt	cccactccctg	atttctggaa	aagagctagg	aaggacaggc	1080
aacttggcaa	atcaaaggccc	tgggactagg	gggttaaaat	acagctccc	ctctcccac	1140
ccggccccagt	ctctgtccct	tttggtaggag	ggacttagag	aagggggtgg	cttgcctgt	1200
ccagtttaatt	tctgaccttt	actccctgccc	tttgagtttgc	atgatgtca	gtgtacaagc	1260
gttttctccc	taaagggtgc	agctgagcta	ggcagcagca	agcatccctg	gggtggcata	1320
gtgggggtgt	gaataccatg	tacaaagctt	gtgcccagac	tgtgggtggc	agtgcggccac	1380
atggccgctt	ctccttggaa	ggcttcgtat	gactgggggt	gttgggcagc	cctggaggct	1440
tcagttgcag	ccatgcctta	agccaggcca	gcctggcagg	gaagctcaag	ggagataaaa	1500
ttcaacctct	tggggccctcc	tgggggtaa	gagatgtgc	attcgcctc	ttaatggga	1560
ggtggccctag	ggctgctcac	atattctgga	ggagccctcc	ctcctcatgc	cttcttgcc	1620

-continued

cttgcgtctt ag

1632

<210> SEQ ID NO 10
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(100)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(100)
<223> OTHER INFORMATION: This sequence may encompass 60-100 nucleotides

<400> SEQUENCE: 10

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	100

<210> SEQ ID NO 11
<211> LENGTH: 1845
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 11

atgccccgggt tttacgagat tgtgattaag gtcccccagcg accttgacga gcatctgcc	60
ggcatttctg acagcttgcgt gaaactgggtg gccgagaagg aatggggagtt gccgcggcagat	120
tctgacatgg atctgaatct gattgagcag gcacccctga ccgtggccga gaagtcgcag	180
cgcgcacttc tgacggaaatg ggcgcgtgtg agtaaggcccc cggaggccct tttctttgt	240
caatttgaga agggagagag ctacttccac atgcacgtgc tcgtggaaac caccggggtg	300
aaatccatgg ttttgggacg tttcttgagt cagatccgc aaaaactgtat tcagagaatt	360
taccgcggga tcgagccgac tttgcacaaac tgggttcgcgg tcacaaagac cagaatggc	420
gcccggaggcg ggaacaaggt ggtggatgag tgcatacatcc ccaattactt gctccccaaa	480
acccagectg agctccagt ggcgtggact aatatggaaac agtatttaag cgccgtttg	540
aatctcacgg agcgttaaacg gttgggtggc cagcatctga cgcacgtgtc gcagacgcag	600
gagcagaaca aagagaatca gaatccaaat tctgtatgcgc cgggtatcag atcaaaaact	660
tcagccaggt acatggagct ggtgggtgg ctcgtggaca aggggattac ctcggagaag	720
cagtggatcc aggaaaatca ggagagctac ctctccttca actccacccgg caactctcg	780
agccagatca aggccgcgct cgacaacgcg accaaaattha tgagtctgac aaaaagcgcg	840
gtggactacc tcgtggggag ctccgttccc gaggacattt caaaaaacag aatctggcaa	900
atttttgaga tgaatggcta cgacccggcc tacgcgggtt ccattctcta cggctgggt	960
cagcgctct tcaacaagag gaacaccgtc tggctctacg gaccggccac gaccggcaag	1020
accaacatcg cggaggccat cgccccacact gtgcctttt acggctgcgt gaactggacc	1080
aatgaaaact ttcccttaa tgactgtgt gacaaaatgc tcattttgggt ggaggaggaa	1140
aagatgacca acaagggttgt tgaatccgccc aaggccatcc tggggggctc aaagggtgcgg	1200
gtcgatcaga aatgtaaatc ctctgttcaa attgattcta cccctgtcat tgtaacttcc	1260
aatacaaaaca tgtgtgttgtt ggtggatggg aattccacga cctttgaaca ccagcagccg	1320
ctggaggacc gcatgttcaa atttgaactg actaagcggc tcccgccaga ttttggcaag	1380

-continued

attactaagc aggaagtcaa ggacttttt gcttggcaa aggtcaatca ggtgcggtg	1440
actcacgagt ttaaagttcc cagggaaattt gcgggaaacta aaggggcgga gaaatctcta	1500
aaacgcccac tgggtgacgt caccataact agctataaaa gtctggagaa gcgggccagg	1560
ctctcatttg ttccccagac gcctcgcagt tcagacgtga ctgttgatcc cgctctctg	1620
cgcacgcgtca attggaaattt aaggatgtat tgcaaatgtg actatcatgc tcaatttgac	1680
aacatttctta acaaattgtga tgaatgtgaa tatttgaatc ggggcaaaaa tggatgtatc	1740
tgtcacaatgt taactcaactg tcaaatttgtt catggattt cccctggga aaaggaaaac	1800
ttgtcagatt ttggggattt tgacgatgcc aataaagaac agtaa	1845

<210> SEQ ID NO 12

<211> LENGTH: 1866

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 12

atgcgggggtt ttacgagat tgtgattaag gtccccagcg accttgcgca gcatctgccc	60
ggcatttctg acagcttgcgtt gaaactgggtg gccgagaagg aatggggat ttccggccagat	120
tctgacatgg atctgaatctt gattgagcg gcacccctga ccgtggccga gaagctgcag	180
cgcgacttttc tgacggaatg gcgccgtgtg agtaaggccc cggaggccct tttctttgt	240
caatttgaga aggggagagag ctacttccac atgcacgtgc tcgtggaaac caccgggtg	300
aaatccatgg ttttgggacg tttctgtt cagattcgcg aaaaactgtat tcagagaatt	360
taccgcggga tcgagccgac tttgcacaaac tggttcgcgg tcacaaagac cagaatggc	420
gccggaggcg ggaacaagggt ggtggatgag tgctacatcc ccaatttactt gctccccaaa	480
acccagcctg acgtccagtg ggcgtggact aatatggaaac agtattttaa cgcctgtttt	540
aatctcacgg agcgtaaacg gttgggtggcg cagcatctga cgcacgtgtc gcagacgcag	600
gagcagaaca aagagaatca gaatcccaat tctgtatgcgc cgggtatcag atcaaaaact	660
tcagccaggt acatggagct ggtcggttgg ctcgtggaca aggggattac ctcggagaag	720
cagtggatcc aggaggacca ggcctcatac atctccttca atgcggccctc caactcgcgg	780
tcccaaatac aggctgcctt ggacaatgcg gggaaatgtt tgagcctgac taaaaccgcc	840
cccgactacc tgggtggccca gcagccgtg gaggacattt ccagcaatcg gattttataaa	900
attttggaaac taaacgggta cgatccccaa tatgcggott ccgtctttctt gggatggcc	960
acgaaaaaaatg tcggcaagag gaacaccatc tggctgtttt ggcctgcaac taccggaaag	1020
accaacatcg cggaggccat agcccacact gtgccttctt acgggtgcgt aaactggacc	1080
aatggaaact ttcccttcaa cgactgtgtc gacaagatgg tgatctggtg ggaggagggg	1140
aagatgaccg ccaagggtcggt ggagtggccaa aaagccattt ctcggaggaag caaggtgcgc	1200
gtggaccaga aatgcagtc ctcggcccaat atagaccggc ctcccgatcg cgtcacctcc	1260
aacaccaaca tgcgtgcgtt gatttgcggg aactcaacga ctttcgaaca ccagcaggcc	1320
ttgcggaccacc ggtgttccaa attttgcgtt accccggccgt tggatcatga ctttggaaag	1380
gtcaccatcg aggaaggtaaa agactttttc cgggtggccaa aggtatcgat ggttggatgt	1440
gagcatgaat tctacgtcaa aaagggtgga gccaagaaaa gacccggcccc cagtgcgc	1500
gatataatgtt agcccaaaacg ggtgcgcgcg tcaatgcgc acggatcgac gtcagacgcg	1560
gaagcttcga tcaactacgc agacaggtac caaaacaaat gttctcgatca cgtggccatg	1620
aatctgtatgc tggggatccctg cagacaatgc gagagaatga atcagaatcc aatatctgc	1680

-continued

ttcaactcacg gacagaaaaga ctgttttagag tgcttcccg tgcagaatc tcaaccgtt	1740
tctgtcgta aaaaggcgta tcagaaaactg tgctacattc atcatatcat gggaaaggtg	1800
ccagacgctt gcactgcctg cgatctggtc aatgtggatt tggatgactg catcttgaa	1860
caataa	1866
<210> SEQ ID NO 13	
<211> LENGTH: 2211	
<212> TYPE: DNA	
<213> ORGANISM: Adeno-associated virus	
<400> SEQUENCE: 13	
atggctgccg atggtttatct tccagattgg ctcgaggaca accttagtga aggaattcgc	60
gagtgggtggg ctttgaacc tggagccccct caacccaagg caaatcaaca acatcaagac	120
aacgcgtcgag gtcttgtgct tccgggttac aaataccctg gacccggcaa cggactcgac	180
aaggggggagc cggtaacgcg acgcagacgcg gccccctcg agcacgacaa ggcctacgac	240
cagcagctca aggccggaga caacccgtac ctcaagtaca accacgocga cgccgagttc	300
caggagccggc tcaaagaaga tacgtctttt gggggcaacc tggggcggagc agtcttccag	360
gccaaaaaga ggcttcttga accttttggt ctgggttggg aagcggotaa gacggctct	420
ggaaagaaga ggcctgtaga gcagtctct caggaaccgg actcctccgc gggatttggc	480
aaatcgggtg cacagccgc taaaaagaga ctcaatttcg gtcagactgg cgacacagag	540
tcagtcccag accctcaacc aatcgagaa cttcccgacccccctcagg tgtggatct	600
cttacaatgg cttcaggtgg tggegcacca gtggcagaca ataacgaagg tgccgatgga	660
gtgggttagt cctcgaaa ttggcattgc gattccaaat ggctggggga cagagtcatc	720
accaccagca cccgaacctg ggcctgccc acctacaaca atcacctcta caagcaatc	780
tccaacagca catctggagg atcttcaat gacaacgcct acttcggcta cagcaccccc	840
tgggggtatt ttgactcaa cagattccac tgccacttcc caccacgtga ctggcagcga	900
ctcatcaaca acaactgggg attcggct aagcgactca acttcaagct cttcaacatt	960
caggtcaaag aggttacgga caacaatggc gtcaagacca tgcacccataa cttaccagc	1020
acggtccagg tttcacggc ctcagactat cagttcccg acgtgtcggt gtcggctcac	1080
gagggctgcc tcccgccgtt cccagcgac gtttcatga ttccctagta cgggtatctg	1140
acgcttaatg attggaaagcca ggcgtgggt cgttcgtctt tttactgcct ggaatatttc	1200
ccgtcgcaaa tgctaagaac gggtaacaac ttccagttca gctacgagtt tgagaacgta	1260
cctttccata gcagctacgc tcacagccaa agcctggacc gactaatgaa tccactcatc	1320
gaccaatact tttactatct ctcaaagact attaacgggtt ctggacagaaa tcaacaaacg	1380
ctaaaattca gtgtggccgg acccagcaac atggctgtcc agggaaagaaa ctacataacct	1440
ggacccagct accgacaaca acgtgtctca accactgtga ctcaaaacaa caacagcgaa	1500
tttgcttggc ctggagcttcc ttcttggctt ctcaatggac gtaatagctt gatgaatcct	1560
ggacctgctta tggccagcca caaagaagga gaggaccgtt ttttccttt gtctggatct	1620
ttaatttttgc gcaaacaagg aacttggaaaga gacaacgtgg atgcggacaa agtcatgata	1680
accaacgaag aagaaattaa aactactaac ccggtagcaa cggagtccta tggacaagtg	1740
gccacaaaacc accagagtgc ccaaggcacag ggcagacccg gctgggttca aaaccaagga	1800
atacttccgg gtatggtttgc gaggacaga gatgtgtacc tgcaaggacc catttggcc	1860

-continued

aaaattcctc acacggacgg caactttcac ccttctccgc tcatggagg gtttgaatg	1920
aagcaccgcg ctcctcagat cctcatcaaa aacacacctg tacctcgcca tcctccaacg	1980
gccttcaaca aggacaagct gaactcttc atcacccagt attctactgg ccaagtcaagc	2040
gtggagatcg agtgggagct gcagaaggaa aacagcaagc gctggaaccc ggagatccag	2100
tacacttcca actattacaa gtctaataat gttgaatttg ctgttaatac tgaagggtgta	2160
tatagtgAAC cccgccccat tggcaccaga tacctgactc gtaatctgta a	2211

<210> SEQ ID NO 14
<211> LENGTH: 12003
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 14

aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaacga	60
cgcgcgcgta actcacgtta agggattttg gtcatgagct tgccgcgtcc cgtcaagtca	120
gcgtaatgtc ctgcttttag aaaaactcat cgagcatcaa atgaaactgc aatttattca	180
tatcaggatt atcaataccca tattttgaa aaagccgtt ctgtaatgaa ggagaaaaact	240
caccgaggca gttccatagg atggcaagat cctggatcgt gtctgcgtt ccgactcgctc	300
caacatcaat acaaccttatt aattccctt cgtcaaaaat aaggttatca agtgagaaat	360
caccatgagt gacgactgaa tccggtgaga atggcaaaaat tttatgcatt tctttccaga	420
cttggtaaac aggccagcca ttacgctcgat catcaaaatc actcgcatca accaaaccgt	480
tattcattcg tgattgcgcc tgagcgaggg aaaaatcgcg atcgcgttta aaaggacaat	540
tacaaacagg aatcgagtgc aaccggcgca ggaacactgc cagcgcata acaatattt	600
cacctgaatc aggatattct tctaataccct ggaacgctgt tttccgggg atcgcagtgg	660
tgagtaacca tgcatacatca ggagtaacggaa taaaatgtt gatggcgaa agtggataaa	720
attccgtcag ccagttagt ctgaccatct catctgtaa atcattggca acgctacatt	780
tgccatgtt cagaaacaac tctggcgcat cgggcttccc atacaagcga tagattgtcg	840
cacctgattt cccgacatta tcgegagccc atttataccca atataaatca gcatccatgt	900
tggaaattaa tcggggcctc gacgttccc gttgaatatg gtcataatc ttcccttttc	960
aatattttt aagcattttt cagggttattt gtctcatgag cggataacata tttgaatgtt	1020
tttagaaaaaa taaacaaataa ggggtcagtgtt ttacaaccaa ttaaccaattt ctgaacat	1080
tcgcgagccc atttataccctt gaatatggct cataacaccc cttgtttgcc tggcgccagt	1140
agcgcgggtgg tcccacctga ccccatgcgg aactcagaag tgaaacgcgg tagcgccat	1200
ggtagtgtgg ggactccccca tgcgagagta gggaaactgcg aggcatcaaa taaaacgaaa	1260
ggctcagtcg aaagactggg ctttcgccc gggctaattt ggggggtgtcg cccttcgtg	1320
aagggggtgcg cggatccgtta tacgtttaaa cgcggccgcg gtacccaactt ccatgctca	1380
cagtcccccggatgtacccca ccctgcgtcg caaccaggaa cagctctaca gcttctggaa	1440
gcgcactcg ccctacttcc gcagccacag tgccgcaggat aggagcgcac cttcttttg	1500
tcacttgaaa aacatgtaaa aataatgtac tagagacact ttcaataaaag gcaaatgttt	1560
ttatgggttac actctcggtt gattattac cccccccctt gccgtctgcg ccgtttaaaa	1620
atcaaagggg ttctgcgcgcg catcgctatg cgccactggc agggacacgt tgcgataactg	1680

US 12,383,587 B2

59**60**

-continued

gtgttttagtgc	ctccacttaaa	actcaggcac	aaccatccgc	ggcagctcggt	tgaagtttc	1740
actccacagg	ctgcgcacca	tcaccaacgc	gtttagcagg	tggggcgccg	atatcttgc	1800
gtcgcagttg	gggcctccgc	cctgcgcg	cgagttgcga	tacacaggg	tgcagcactg	1860
gaacactatac	agcgcgggt	ggtgcacgt	ggccagcacg	ctttgtcg	agatcagatc	1920
cgcgtccagg	tcctccgcgt	tgctcagg	gaacggagt	aacttggta	gctgcctcc	1980
caaaaagggc	cgctgeccag	gcttgagtt	gcactgcac	cgtagtg	tcaaaagg	2040
accgtgeccg	gtctggcg	taggatacag	cgctgcata	aaagcctg	tctgtttaaa	2100
agccacatgt	gccttgcgc	cttcagagaa	gaacatgc	caagacttgc	cggaaaactg	2160
attggccgga	caggccgcgt	cgtgcacg	gcaccttgc	tgggtgttgg	agatctgcac	2220
cacatttcg	ccccacccgt	tcttcacgt	cttgccttgc	ctagactgt	cttcagcgc	2280
gctgtgc	tttgcgtcg	tcacatccat	ttcaatcag	tgccttttat	ttatcataat	2340
gcttcgtgt	agacacttaa	gctgccttc	gatctcag	cagcgg	gccacaacgc	2400
gcagccgt	ggctcgtgt	gcttgtaggt	cacctctgc	aacgactg	ggta	2460
caggaatcgc	cccatcatcg	tcacaaagg	cttgcgtcg	gtgaagg	tgcac	2520
cggtgtctcc	tcgttcag	aggcttgc	tacggccgc	agagcttcca	cttgg	2580
cagtagtttgc	aagttgc	ttagatcg	atccacgt	tacttgc	tcagcgc	2640
cgcagcctcc	atgccttc	cccacgc	acgc	actcag	ggttc	2700
cgtaatttca	cttccgc	cgctgg	ttcc	tcttgc	gcata	2760
cgcactgg	tcgttc	tcagcc	cactgtgc	ttac	cttgc	2820
gattagcacc	ggtgggttgc	tgaaaccc	catttgc	gccacat	cttgc	2880
ctcgctgtcc	acgattac	ctgg	cg	ggcttgg	aagg	2940
cttttcttc	ttgggc	caa	cgcc	ggag	gtcgat	3000
tgtgcggc	accagcgc	cttgc	gtctc	tcctcg	cgata	3060
cctcatccgc	tttttgg	gcccgg	aggcgg	gacgg	gggac	3120
gtcctccatg	gttgggg	ac	ccgcgt	cg	ctcg	3180
ctgctctct	tccgact	ccat	ctc	tatagg	caga	3240
agtcgagaag	aagg	ac	ct	tcgt	gact	3300
tgccgcaac	gcgc	ctacca	c	cg	gttgg	3360
gattatcg	cagg	accc	gtt	gg	agg	3420
agaggataaa	aag	caag	agg	acaac	gag	3480
ggacgaaagg	catgg	gact	ac	tagatgt	ggg	3540
gcccgcgt	gc	cattat	gc	gag	gc	3600
agcggatgtc	ac	cttgc	ac	tc	cc	3660
ccaagaaaac	ggc	acatgc	ag	ccaa	cc	3720
gccagagg	cttgc	cc	atc	tc	cc	3780
ccgtgc	cc	gc	ac	gg	ct	3840
tgatatcg	tcg	ctca	a	at	tt	3900
gcccgcgg	aa	cg	ct	gg	at	3960
ggtggaa	actc	gagg	tg	ac	tc	4020
ccactttg	cc	ac	cc	gg	ac	4080

-continued

gctgatcggtc cgccgtgcgc agccccctgga gagggatgca aatttgcaga aacaaacaga 4140
 ggaggggccta cccgcgtttc gcgacgagca gctagcgcgc tggcttcataa cgcgcgagcc 4200
 tgccgacttg gaggagcgac gcaaactaat gatggccgca gtgctcgta ccgtggagct 4260
 tgagtgcgtc cagcggttct ttgctgaccc ggagatcgac cgcaagctag agggaaacatt 4320
 gcactacacc tttcgacagg gctacgtacg ccaggcctgc aagatctcca acgtggagct 4380
 ctgcaacactg gtctcttacc ttgaaatttt gcacgaaaac cgccttgggc aaaacgtgct 4440
 tcattccacg ctcaaggcg aggccgcgcg cgactacgtc cgcaactgct tttacttatt 4500
 tctatgtcac acctggcaga cggccatggg cgtttggcag cagtgttgg aggagtgcaa 4560
 cctcaaggag ctgcagaaac tgctaaagca aaacttgaag gacctatgga cggccttcaa 4620
 cgagcgtccc gtggccgcgc acctggcggaa catcatttt cccgaacgcc tgcttaaac 4680
 cctgcaacag ggtctgccag acttcaccag tcaaagcatg ttgcagaact ttaggaactt 4740
 tatacttagag cgctcaggaa tcttgcgcgc cacctgtgtc gcaacttccat gcaactttgt 4800
 gccccatataa taccgcgaat gccctccgcgc gctttggggc cactgttacc ttctgcagct 4860
 agccaaactac cttgcctacc actctgtacat aatggaaagac gtgagcgggtg acggtctact 4920
 ggagtgtcac tgctcgctgca acctatgca cccgcacccgc tccctgggtt gcaattcgca 4980
 gctgcttaac gaaagtcaaa ttatcggtac ctttggatgtc cagggtccct cgcctgacga 5040
 aaagtcggcg gctccggggt tgaaactcac tccgggggtt tggacgtcgg cttacctcg 5100
 caaaatgttca cctgaggact accacgcaca cgagattagg ttctacgaag accaatcccg 5160
 cccgcaccaat gcgaggctt ccgcctgcgt cattaccag ggccacattt ttggccaatt 5220
 gcaaggccatc aacaaagccc gccaagagtt tctgctacga aaggacggg gggttactt 5280
 ggaccccccag tccggcgagg agctcaaccc aatccccccg cccgcgcgc cctatcagca 5340
 gcagccgcgg gcccttgctt cccaggatgg caccaaaaaa gaagctgcag ctgcgcgcgc 5400
 caccacccgaa cgaggaggaa tactggaca gtcaggcaga ggagggtttt gacgaggagg 5460
 aggaggacat gatggaaagac tgggagagcc tagacggagaa agcttccgag gtcgaagagg 5520
 tgctcagacga aacaccgtca ccctcggtcg cattccctc gccggcgccc cagaaatcg 5580
 caaccgggttc cagcatggct acaacctccg ctctcgaggc gccgcggca ctggccgttc 5640
 gcccacccaa ccgtatgtt gacaccactg gaaccaggcc cggtaagtcc aagcagccgc 5700
 cggccgttgc ccaagagcaa caacagcgcc aaggctaccg ctcatggcgc gggcacaaga 5760
 acgccatagt tgcttgcttca aacactgtt gggcaacat ctccctcgcc cggccgttcc 5820
 ttctctacca tcacggcggtt gccttccccc gtaacatctc gcattactac cgtcatctc 5880
 acagccataa ctgcaccggc ggcagcggca gccggcagca cagcagcgcc cacacagaag 5940
 caaaaggcgac cggatagcaa gactctgaca aagcccaaga aatccacagc ggccggcagca 6000
 gcaggaggag gagcgtcgcc tctggcgccc aacgaacccg tatcgacccg cgagctttaga 6060
 aacaggattt ttcccaacttct gtatgtata ttcaacaga gcaaggccca agaacaagag 6120
 ctgaaaataa aaaacagggtc tctgogatcc ctacccgcgca gtcgcgttca cacaacaaagc 6180
 gaagatcagc ttccggcgcac gctggaaagac gccggaggctc tcttcgttata atactgcgcg 6240
 ctgacttta aggacttagt tcgcgcctt tctcaattt aagcgcgaaa actacgtcat 6300
 ctccagcggc cacacccggc gccagcacct gtcgtcagcg ccattatgag caaggaaatt 6360
 cccacgcctt acatgtggag ttaccagcca caaatgggac ttgcggctgg agctgcccac 6420

US 12,383,587 B2

63**64**

-continued

gactactcaa cccgaataaa ctacatgago gcgggacccc acatgatatac ccgggtcaac	6480
ggaatccgcg cccaccgaaa ccgaaattctc ttggAACAGG cggctattac caccacacct	6540
cgtataacc ttaatccccg tagttggccc gctgccctgg tgtaccagga aagtcccgt	6600
cccaccactg tggtacttcc cagagacgccc caggccgaag ttcatgac taactcaggg	6660
gcccggctt tcgtcacagg gtgcggcgcc cgccggacgc taggttttag	6720
ggccggatggaa cttgtatgtt ttggaaattt tagttttctt aaaatgggaa gttacgtaac	6780
gtggggaaac ggaagtgcacg atttgaggaa gttgtgggtt ttttggctt cgtttctggg	6840
cgtaggctcg cgtcgccgtt tctgggtttt ttttgcggac tttaaccgtt acgtcattt	6900
ttagtcctat atatactcgc tctgcacttgc gcccccccc actctgtac tgattgagct	6960
ggtgcgtgtt cggatgggtt ttttttaata ggttttttt tttactggta aggctgactg	7020
ttatggctgc cgctgtggaa ggcgtgtatg ttgttctggaa gcccgggggt gctattttgc	7080
ctaggcagga gggttttca ggtgtttatg ttgttcttc tcctattaat tttgttatac	7140
ctccctatggg ggctgtatg ttgtctctac gcctgcgggtt atgtattccc ccgggttatt	7200
tcggtcgtt ttagcactg accgatgtac atcaacactga tttgtttacc gagtttaca	7260
ttatgactcc ggacatgacc gaggagctgt cgggtgggtt ttttacac ggtgaccagt	7320
tttttacgg tcacgcggc atggccgttgc tccgtcttat gcttataagg gttgttttc	7380
ctgttgcgtt aacaggcttctt aatgtttaaa tttttttttt ttatttttt ttgtgtttat	7440
gcagaaaccc gcagacatgt ttgagagaaa aatgggtgtt ttttctgtgg tggttccggaa	7500
gcttacctgc ctttatctgc atgagcatga ctacgtgttgc tttttttt tgccgtggc	7560
tttgcctgtt ttttgcgtt gcacccatgc ttttatatcg ccggccatgc aacaagctt	7620
catcggggctt acgctgggtt gcatagctcc gagtatgtgtt gtcataatca gtgtgggtt	7680
ttttgtcatg ttccctggcg gggaaatggc cgcgtggc cgtgcagacc tgacgttac	7740
ttttcagctg gcccctgcgaa gggacctacg ggatcgggtt atttttgtt atgttccgtt	7800
tttgaatctt atacaggctt gtgaggaacc tgaatttttgc caatcatgtat tcgtgttttgc	7860
aggctgttgcg tggaggccgc tctggagcgtt atttttacaa tggccggact taatattcgg	7920
gatttgcgtt aagatatttttgc gagaagggtgg cgagatgaga attttttggg catgggttgc	7980
gggtgcgtggaa tttttataga ggatccac cctgtttttt ttagccctt cgtccacttgc	8040
gacgtgagggg ccgttgcctt tttggaaagcc atttgtcaac atcttacaaa tgccattatc	8100
tgttctttgg ctgttagatgt tgaccacgc accggagggg agcgcgttca cttaatagat	8160
cttcatttttgcgtt aggttttgcgtt taatcttttgcgtt gaataaaaaaaa aaaaacatgg ttcttccagc	8220
tcttcccgctt cttcccgctt gtgactcgca gaacgtatgtt gttaggttgc tgggtgtggc	8280
ttattctgcgtt gtgggtggatg ttatcaggcc accggcgcat gaaggagttt acatagaacc	8340
cgaagccagg gggccgttgcgtt atgttttgcgtt aggttgcgtt tactacaact actacacaga	8400
ggcgtatctaag cggcgagacc ggagacgcgtt acgtgtttgcgtt caccggccgtt cctgggttttgc	8460
cttcaggaaat tatgactacg tccggcgatc catttggcat gacactacga ccaacacgtt	8520
ctcggttgcgtt tcggcgact ccgtacagta gggatcgttgcgtt acctccctttt gagacagaaa	8580
cccgccgttac catactggag gatcatccgc tgctgcccgtt atgttacact ttgacaatgc	8640
acaacgttgcgtt ttagtgcgtt ggtttccctt gcaatgtgggg attttacgtt atttgcgtt	8700
gggttgcgttcc ctgggtatgtt gttctaacgc gggaggagct tggtaatccgtt aggaagtgtt	8760
tgcacgtgttgcgtt cctgtgttgcgtt gccaacatttgcgtt atatcatgtac gaggatgttgcgtt atccatgtt	8820

-continued

acgagtcctg ggctctccac tgtcattgtt ccagtccgg ttcctgcag tgtatagccg 8880
 gcgggcaggc tttggccagc tggtttagga tgggtggaa tggcgccatg ttatcaga 8940
 gtttatatg gtaccgggag gtggtaatt acaacatgcc aaaagaggta atgttatgt 9000
 ccagcgttt tatgaggggt cgccacttaa tctacctgcg cttgtggat gatggccacg 9060
 tggttctgt ggtcccccgc atgagcttg gatacagcgc cttgcactgt gggatttga 9120
 acaatattgt ggtgtgtgc tgcaacttgcgtt aagttagatc aggggtgcgt 9180
 gtgtgcccggaggacaagg cgccattatgc tgccggcggt gcaatcgtc gctgaggaga 9240
 ccactgcatt gttgttattcc tgcaaggacgg agccggcggt gcaacgtttt attcgccgc 9300
 tgctgcagca ccaccgcctt atcctgatgc acgattatga ctctaccccc atgttaggcgt 9360
 ggacttctcc ttgcggccccc gttaaagcaac cgcaagttgg acagcgcct gtggctcgc 9420
 agctggacag cgacatgaac ttaagtggc tgccggggaa gttttaat atcaatgt 9480
 agcgtttggc tcgacaggaa accgtgtggaa atataaacacc taagaatatg tctgttaccc 9540
 atgatatgtat gcttttaag gccagccggg gagaaaggac tgtgtactct gtgtgttggg 9600
 agggagggtgg cagggtgaat actagggttc tgtgagttt attaaggatc ggtgtctgt 9660
 ataagctatg tggtgttggg gctataactac tgaatgaaaa atgacttggaa attttctgca 9720
 attgaaaat aaacacgttg aaacataaca caaacgattt tttattctt ggcaatgtat 9780
 gaaaaaagtgt aagaggatgt ggcaaataatt tcattaatgt agttgtggcc agaccagtcc 9840
 catgaaaatg acatagagta tgcaattggaa gttgtgtctc ctgtttctg ttttccgttt 9900
 agtgttaattt aaccggggat gtcggccctc ctgacgggtt aggaggaggg gaggggtggcc 9960
 tgcattgtctc ccgtgtctt tgcttttgcgctt gctgttggagg agggggggcgatc atctggccca 10020
 gcacccggatg catctggaa aagaaaaaaa ggggctcgcc cctgtttccg gagaaattt 10080
 caagcggggtt ctgcattgac gggggaggaa accccgttc gcccgcgttcc ggcgggtcc 10140
 agactcgaaac cgggggtccc gcgactcaac ccttggaaaa taaccctccg gctacaggaa 10200
 gggggccact taatgttttgcgtt cttccago ctaaccgtt acgctgcgcg cggccagtgg 10260
 ccaaaaaagc tagcgcagca gcccgcgcg ctggaaaggaa gccaaggaa gcactcccc 10320
 gttgtctgac gtcgcacacc tgggttcgac acggggggggtaaccgcattt gatcaggcg 10380
 gacggccgga tacggggctc gaacccgggtt cgtccgcattt gatacccttgc gcaatttac 10440
 caccagacca cggaaagatc cccgttaca ggtctcttgc ttgcacggcc tagacgtca 10500
 acgattgcgc ggcgttgacc ggccagagecg tcccgaccat ggagcacttt ttgcggctgc 10560
 gcaacatctg gaaccgcgtt cgcgttttc cgcgcgcctt caccaccggcc gccggcatca 10620
 cctggatgtc caggttgcattt tacggatattt attaattttt gttacttgcg cgcgtacaa 10680
 acgtcaaaag ggcgcacacaa aatttttattt aatgcataaa taaatactga taacatctta 10740
 tagtttgtat tatattttgtt attatcgttgcgtt acatgtatc ttttgcattt aaaaactgtat 10800
 ttcccttta ttatccgtt gatttattttt ctttatttgcgtt ttacaaacatc tagaaatatt 10860
 gtatatacaa aaaatataaa ataatagatg aatagtttaa ttataggtgtt tcatcaatcg 10920
 aaaaagcaac gtatccattt taaaatgcgt tggcttttgcgtt tcatttataa ggttaatataa 10980
 ttctcatata tcaaggcaac tgcacaggccg ctttataat tctgacaaat gtccttccc 11040
 taaaactcccccc cccataaaaaaa accccggccaa gccgggtttt acgttttttg cggatataacg 11100
 attactcgatgtt atcagaaccg cccaggggccgcg cgcgttttttgcgtt acgttttttg cggatataacg 11160

-continued

cacagaaaaga gttttagaa acgaaaaag gccatccgtc aggggccttc tgcttagtt 11220
 gatgcctggc agttccctac tctcgcccttc cgttccctcg ctcaactgact cgctgcgtc 11280
 ggtcgttcgg ctgcggcgag cggtatcago tcactcaaag gcgtaatac gttatccac 11340
 agaatcaggg gataacgcag gaaagaacat gtgagaaaaa ggccagaaaa aggccaggaa 11400
 ccgtaaaaag ggcgcgttgc tggegtttt ccataggctc cgccccctg acgagcatca 11460
 caaaaatcga cgctcaagtc agagggtggc aaacccgaca ggactataaa gataccaggc 11520
 gttccccctt ggaagetccc tcgtgcgtc ttctgttccg accctgcgc ttaccggata 11580
 cctgtccggc tttctccctt cgggaagcgt ggcgtttt catagtcac gctgttagta 11640
 ttcagttcg gtgttaggtcg ttgcgtccaa gctgggtgt gtgcacgaac ccccggtca 11700
 gcccggccgc tgccgttataa ccggtaacta tcgtctttag tccaacccgg taagacacga 11760
 cttatcgcca ctggcagcag ccactggtaa caggattgc agagcgaggt atgtaggcgg 11820
 tgctacagag ttcttgaagt ggtgggctaa ctacggctac actagaagaa cagtatttg 11880
 tatctgcgt ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg 11940
 caaacaaacc accgctggta gcggtggtt tttgtttgc aagcagcaga ttacgcgcag 12000
 aaa 12003

<210> SEQ ID NO 15
 <211> LENGTH: 17817
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 15

ttagaaaaac tcatcgagca tcaaatttgc ctgcaatttta ttcatatcg gattatcaat 60
 accatattttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga ggcagttcca 120
 taggatggca agatcctggt atcggctctgc gattccgact cgtccaaatcat caatacaacc 180
 tattaatttc ccctcgtaaa aaataagggtt atcaagttagt aaatcaccat ggtgacgac 240
 tgaatccgggt gagaatggca aaagttttagt catttcttgc cagacttggt caacaggcca 300
 gcccattacgc tcgtcatcaa aatcaactcgat atcaacccaa ccgttattca ttctgtattg 360
 cgcctgagcg aggcgaaata cgcgatcgat gttaaaagga caattacaaa caggaatcga 420
 gtgcacccggc cgcaggaaaca ctgcaggcgc atcaacaata ttttccatcg aatcaggata 480
 ttcttctaattt accttggaaacg ctgttttcc ggggatcgca gtggtagt accatgcattc 540
 atcaggagta cggataaaat gcttggatgtt cggaaatggc ataaattccg tcagccagg 600
 tagtctgacc atctcatctg taacatcattt ggcaacgcta ctttgcattt gtttcaaaa 660
 caactctggc gcategggtc tccctatcaa gcgtatcgat gtgcacccgt attgcccac 720
 attatcgca gcccattttt accatataaa atcagcatcc atgttggat ttaatcgccg 780
 cctcgacgtt tccctgttgc tatggctcat attttccattt ttcaatattt attgaagcat 840
 ttatcagggtt tattgtctca tgagggata catatggaa tggatggataaaataaaca 900
 aataggggtc agtggatcaa ccaatttacca aattctgaaat attatcgca gcccatttt 960
 acctgaatggc ggctcataac acccccttgc tgcctggccg cagtagcgcg gtggccac 1020
 ctgaccccat gcccggccactca gaagtggaaac gcccgttagcgc cgtatggtagt gtggggactc 1080
 cccatgcggag agtagggaaac tgccaggcat caaataaaac gaaaggctca gtcgaaagac 1140

-continued

ttggcccttc gccccggcata attagggggt gtgcgccttc gctgagggg tgaccgtata 1200
cgtttaact taattaatcg acagaagcac catgtccttgc ggtcggccct gctgaatcg 1260
caggcggtcg gccatgccccc aggctcggtt ttgacatcgcc cgcaaggcttt tgtagtagtc 1320
ttgcatttgcgc ctttctaccg gcacttcttc ttctccttcc tcttgccttg catctttgc 1380
atctatcgct gggggggcgcc cgaggttgg ccgttaggtgg cgcccttcc ctccatcg 1440
tgtgaccctcg aagccctca tcggctgaag cagggctagg tcggcgacaa cgccgtcgcc 1500
taatatggcc tgctgcacct gcgtgagggt agacttggaa tcatccatgt ccacaaagcg 1560
gtggatcgcc cccgtgttga tgggtgttaagt gcagttggcc ataacggacc agttaacgg 1620
ctggtgaccgc ggctgcgaga gctcggtgta cctgagacgc gagtaagccc tcgagtc 1680
tacgttagtcg ttgcgaatcc gcaccaggta ctggtatccc accaaaaaggcg cgccggccgg 1740
ctggcggtag agggggccagc gttaggggtgc cggggctccg gggggagat cttccaaat 1800
aaggcgatgatatccgtaga tggatcttgc catccaggtg atgcggccgg cggtgttgc 1860
ggcgccgcca aagtgcgcca cgccgttcca gatgttgcgc agcggaaaaa agtgcctcat 1920
ggtcgggacgc ctctggccgg tcaggcgcgc gcaatcggtt acgtcttaga ccgtgc 1980
ggagagcctg taagcgggca ctctccgtg gtctgggttgc taaattcgca agggatcat 2040
ggcgacgac cgggggttgcg gccccgtatc cggccgtccg ccgtgtatcca tgccgttacc 2100
gccccgtgt cgaaccacagg tggatcgacgt cagacaacgg gggagtgctc ctttggctt 2160
cttccaggc gcggcggtcg ctgcgttgcg tttttggcc actggccgcgc cgccagcgtaa 2220
ggggtaggc tggaaagcga aagcattaag tggctcgctc cctgtagccg gaggttatt 2280
ttccaaagggt tgagtcgcgg gaccccccgtt tggatcgatc gaccggccgg actgcggcga 2340
acgggggtttt gctccccgtt catgcaagac cccgcttgc aattctccg gaaacaggga 2400
cgagccccctt ttttgcggg cccagatgc tccgggtctc cggcagatgc gccccctcc 2460
tcagcagcgg caagagcaag agcagcgca gacatgcagg gcaccctccc ctcccttac 2520
cggtcaggaa gggggacat ccgggttgc cggccgcacat gatgtgttatt acgaaaaaaa 2580
ggggcgccgg gccccggactt acctggactt ggaggagggc gagggcttgg cgccgttgg 2640
agccctctt cctgagcggtt acccaagggtt gcaatcgatc cgtgtatcgc gtgggggtt 2700
cgccgcggc cagaacctgtt ttcgcgaccccg cgagggagag gagccccagg agatgcggga 2760
tcgaaaggtc cacgcaggcgc gcgagctcg gcatggcttgc aatcgccgac ggttgcgtcg 2820
cgaggaggac tttgagcccg acgcgcgaaac cgggattttt cccgcgcgcg cacacgttgc 2880
ggccggccgac ctggtaaccgc catacgagca gacgggtgaac caggagatta actttcaaaa 2940
aagctttaaac aaccacgtgc gtacgcttgc ggcgcgcgag gaggttggcttta taggactgt 3000
gcatctgtgg gactttgttgc ggcgcgttgc gcaaaacccaa aatagcaagc cgctcatggc 3060
gcagctgttc cttatagtcg acgcacagcgg ggcacaacggc gcattcaggat atgcgtgtct 3120
aaacatagta gagcccgagg gccgctggct gctcgatttgc ataaacatcc tgccggat 3180
agtgggtgcg gaggcgcagttt tgacgcttgc tgacaaagggtt gccgcacatca actattccat 3240
gettagecttgc ggcaagttt acgcggccaa gatataccat acccccttacg ttcccataga 3300
caaggaggttgc aagatcgagg ggttctacat ggcgcgttgc ctggatgtgc ttacccatgg 3360
cgacgcacccgttgc ggcgcgttgc gcaacgcgcg catccacaag gccgttgcgc tgccggccgg 3420
gcgcgcgttgc agcgcacccgttgc agctgtatgc cagccctgc gggcccttgg ctggcaccggg 3480
cagccggcgat agagaggccg agtctactt tgacgcttgc gctgaccccttgg ctggcaccggg 3540

-continued

aagccgacgc	gccctggagg	cagctggggc	cggaccctggg	ctggcggtgg	cacccgcgcg	3600
cgcggcaac	gtcggcgccg	tggaggaata	tgacgaggac	gatgagtacg	agccagagga	3660
cggcgagta	taagcggtga	tgtttctgtat	cagatgtgc	aagacgcaac	ggacccggcg	3720
gtgcgggggg	cgcgtcagag	ccagccgtcc	ggcccttaact	ccacggacga	ctggcgccag	3780
gtcatggacc	gcatcatgtc	gctgactgct	cgcaatcccg	acgcgttccg	gcagcagccg	3840
caggccaacc	ggctctccgc	aattctggaa	gcccgtggcc	ccggcgccgc	aaacccccacg	3900
cacgagaagg	tgctggcgat	cgtaaacgcg	ctggccgaaa	acagggccat	ccggccccac	3960
gaggccggcc	tggtctacga	cgcgcgtctt	cagcgcgtgg	ctcgttacaa	cagcggcaac	4020
gtgcagacca	acctggaccg	gctgggtgggg	gatgtgcgcg	aggccgtggc	gcagcgtgag	4080
cgcgcgcgcg	acgcaggccaa	cctgggctcc	atgggtgcac	taaacgcctt	cctgagtaca	4140
cagcccgcca	acgtgcgcgc	gggacaggag	gactacacca	actttgtgag	cgcactgcgg	4200
ctaattggta	ctgagacacc	gc当地gttag	gtgttaccat	ctggccaga	ctattttc	4260
cagaccagta	gacaaggccct	gcagaccgta	aacctgagcc	aggcttcaa	aaacttgacg	4320
gggctgtggg	gggtgcgggc	tcccacagcc	gaccgcgcga	ccgtgtctag	cttgcgtgacg	4380
cccaactcgc	gcctgttgct	gctgtaata	gcgccttca	cgacagatgg	cagcgtgtcc	4440
cgggacacat	accttaggtca	cttgctgaca	ctgttccgcg	aggccatagg	tcaggcgcatt	4500
gtggacgagc	atactttcca	ggagattaca	agtgtcagcc	gcccgtggg	gcaggaggac	4560
acggggcagcc	tggaggcaac	cctaaactac	ctgtgtacca	accggggcga	gaagatcccc	4620
tcgttgcaca	gttgcacc	ctttggcgea	tcccatttctc	cagtaacttt	atgtccatgg	4680
gcccactc	agacctgggc	caaaaccttc	tctacgccaa	ctccgcacac	gcccgttagaca	4740
tgactttga	ggtggatccc	atggacgago	ccacccttct	ttatgtttt	tttgaagtct	4800
ttgacgttgt	ccgtgtgcac	cgccgcacc	cgccgtcat	cgaaaccgtg	tacctgcgca	4860
cgccttctc	gcccggcaac	gccacaacat	aaagaagcaa	gcaacatcaa	caacagctgc	4920
cgccatggc	tccagtgcgc	aggaactgaa	agccattgtc	aaagatctg	gttgcggcc	4980
atattttttgc	ggcacctatg	acaagcgctt	tccaggctt	gttctccac	acaagctcgc	5040
ctgcgcata	gtcaataacgg	ccggtcgcga	gactggggc	gtacacttgg	tggccttgc	5100
ctggaaaccgc	cactaaaaa	catgtacct	cttgagcc	tttgcgtttt	ctgaccagcg	5160
actcaagcag	gtttaccat	tttgcgtacga	gtcacttcttgc	cgccgttagcg	ccattgcattc	5220
ttccccccac	cgctgtataa	cgctggaaaa	gtccacccaa	agcgtacagg	ggcccaactc	5280
ggccgcctgt	ggactattct	gctgcgtt	tctccacgccc	tttgcgtact	ggcccaaac	5340
tcccatggat	cacaacccca	ccatgacact	tattaccgg	gtacccact	ccatgctcaa	5400
cagtccccac	gtacagccca	ccctgcgtcg	caaccaggaa	cagctctaca	gttgcgttgg	5460
gcccactc	ccctacttcc	gcagccacag	tgccgttgc	aggagccca	cttcttttgc	5520
tcacttgaaa	aacatgtaaa	aataatgtac	tagagacact	ttcaataaaag	gcaaatgcctt	5580
ttatggta	actctcggtt	gattattac	ccccaccctt	gcccgttgcgc	ccgtttaaaa	5640
atcaaagggg	ttctgcgcgc	catcgctatg	cgccacttgc	agggacacgt	tgccgtact	5700
gtgttttagtg	ctccacttaa	actcaggac	aaccatccgc	ggcagctcg	tgaagtttc	5760
actccacagg	ctgcgcacca	tcaccaacgc	gtttagcagg	tcggggccgc	atatcttga	5820
gtcgcagttg	gggcctccgc	cctgcgcgcgc	cgagttgcga	tacacagggt	tgcagcact	5880

-continued

gaacactatac agcgccgggt ggtgcacgct ggccagcacg ctcttgcgg agatcagatc	5940
cgcgtccagg tcttcggcgt tgctcagggo gaacggagtc aactttggta gctgcctcc	6000
caaaaagggc gcgtgeccag gcttgagtt gcactgcac cgtagtggca tcaaaagggt	6060
accgtgeccc gtctggcgt taggatacag cgctgcata aaagccttga tctgtttaaa	6120
agccaccta gctttgcgc cttcagagaa gaacatgccg caagacttgc cgaaaaactg	6180
attggccgga caggccgcgt cgtgcacgca gcaccttgcg tcgggtttgg agatctgcac	6240
cacatttccg cccccccggt tcttcacgat cttggccttgc ttagactgtc cttcagcgc	6300
gcgtgeccc tttecgctcg tcacatccat ttcaatcagc tgctccttat ttatcataat	6360
gcttccgtgt agacacttaa gctgccttc gatctcagcg cagcggtgca gccacaacgc	6420
cgagcccggt ggctcggtat gcttgtaggt cacctctgca aacgactgca ggtacgcctg	6480
caggaatcgc cccatcatcg tcacaaagggt cttgttgcgt gtgaagggtca gctgcaaccc	6540
gcgggtgtcc tcgttcagcc aggtcttgca tacggccgccc agagcttcca cttggtcagg	6600
cagtagtttgc aagttgcct ttatgcgtt atccacgtgg tacttgtcca tcagcgcgcg	6660
cgcagcctcc atgccttct cccacgcaga cacgatccgc acactcagcg ggttcatcac	6720
cgttaatttca ctttccgcgtt cgctggcgtt ttcccttcc tcttgcgtcc gcataccacg	6780
cgccactggg tcgtcttcat tcagccgcgc cactgtgcgc ttacctcctt tgccatgctt	6840
gattagcacc ggtgggttgc tgaaaacccac catttgcgtc gccacatctt ctcttccttc	6900
ctcgctgtcc acgattaccc ctggtgatgg cgggcgcgtc ggcttgggag aagggcgttt	6960
ctttttcttc ttgggcgcaa tggccaaatc cgccgcgcgag gtcgatggcc gcgggcttgg	7020
tgtgcgccc accagcgcgt cttgtatga gtcttcctcg tcctcggact cgatacgcgg	7080
cctcatccgc ttttttgggg gcccgggggg aggccggccgc gacggggacg gggacacac	7140
gtcctccatg gttggggac gtcgcgcgc accgcgtccg cgctgggggg tggttcgcg	7200
ctgctccctc tccccactgg ccatttcctt ctccatatagg cagaaaaaga tcatggagtc	7260
agtcgagaag aaggacagcc taaccgcctt ctctgagttc gccaccacccg cctccaccga	7320
tgccgcacac ggcctacca cttccccgt cgaggccacc cccgttggagg aggaggaagt	7380
gattatcgag caggacccag gttttgtaa cgaagacgac gaggaccgt cagtagccac	7440
agaggataaa aagcaagacc aggacaacgc agaggcaaac gaggaacaag tcggggcgccc	7500
ggacgaaagg catggcact acctagatgt gggagacgac gtgctgttga agcatctgca	7560
gcgcgcgtgc gccattatct ggcgcgcgtt gcaagagcgc acgcgtgtgc ccctgcct	7620
agcggatgtc agccttgcct acgaacgcac cctattctca cccgcgtac ccccaaaacg	7680
ccaagaaaac ggcacatgcg agcccaaccc ggcctcaac ttctaccccg tatttgcgt	7740
gccagaggtg cttgccacct atcacatctt tttccaaaac tgcaagatac ccctatcctg	7800
cctgtccaaac cgcagccgag cggacaagca gctggccttgc cggcaggccgctgtcataacc	7860
tgtatatcgcc tcgctcaacg aagtgcacaa aatctttgag ggtcttggac gcgacgagaa	7920
gcgcgcggca aacgcctctgc aacagaaaaa cagcggaaat gaaagtcaact ctggagttt	7980
ggtggactc gaggggtgaca acgcgcgcct agccgtacta aaacgcagca tcgaggtcac	8040
ccactttgcc taccggcac ttaacctacc ccccaagggtc atgagcacag tcatgagtgaa	8100
gctgategtg cgcgcgtgcgc agccctggaa gagggatgca aatttgcac aacaaacaga	8160
ggagggcccta cccgcagttg gcgacgagca gctagcgcgc tggcttcaaa cgcgcgagcc	8220
tgccgacttg gaggagcgcac gcaaactaat gatggccgca gtgctgtta ccgtggagct	8280

-continued

tgagtgcatg cagcggttct ttgctgaccc ggagatgcag cgcaagctag aggaaacatt 8340
 gcactacacc ttgcgacagg gctacgtacg ccaggcctgc aagatctcca acgtggagct 8400
 ctgcaacctg gtctcctacc ttgaaatttt gcacgaaaac cgcctggc aaaacgtgct 8460
 tcattccacg ctcaaggcg aggccgcgcg cgactacgtc cgactgtcg tttacttatt 8520
 tctatgtcac acctggcaga cggccatggg cgttggcag cagtgttgg aggagtgc当地 8580
 cctcaaggag ctgcagaaac tgctaaagca aaacttgaag gacatatgga cggccattcaa 8640
 cgagcgtcc gtggccgcg acctggcga catcatccc cccgaacgcc tgcttaaac 8700
 cctgcaacag ggtctgccag acttcaccag tcaaagcatg ttgcagaact ttaggaactt 8760
 tataccttagag cgctcaggaa tcttgcccgc cacctgtgt gcacttccta gcgactttgt 8820
 gcccattaag taccgcaat gcccctccgc gctttggggc cactgttacc ttctgcagct 8880
 agccaaactac cttgcctacc actctgacat aatggaaagac gtgagcggtg acggctact 8940
 ggagtgtcac tgctcgatc acctatgcac cccgcaccgc tccctgggtt gcaattcgca 9000
 gctgcttaac gaaagtcaaa ttatcggtac ctttgagctg cagggtccct cgcctgacga 9060
 aaagtcccg gctccgggt tgaaaactca tccggggctg tggacgtcgg cttacctcg 9120
 caaatttgtt cctgaggact accacgcccc ctagatttgc ttctacgaa accaatcccg 9180
 cccgc当地aaat gcgagctta ccgcctgcgt cattaccgcg ggccacattc ttggccaaatt 9240
 gcaaggccatc aacaaagccc gccaagagtt tctgctacga aaggacggg gggtttactt 9300
 ggaccccccag tccggcggagg agctcaaccc aatccccccg cccgcgcagc cctatcagca 9360
 gcagccgc当地 gcccctgctt cccaggatgg caccaaaaaa gaagctgcag ctgc当地ccgc 9420
 cacccacggc cggaggaggaa tactggaca gtcaggcaga ggaggtttg gacgaggagg 9480
 aggaggacat gatggaaagac tggagagcc tagacgagga agcttccgag gtc当地agg 9540
 tgctcagacga aacaccgtca ccctcggtcg cattccctc gcggcgeccc cagaaatcg 9600
 caaccgggtc cagcatggct acaacctcg ctctcaggc gccc当地ggca ctgcccgttc 9660
 gccc当地aaat ccgtatgg gacaccactg gaaccaggcc cggtaagtcc aagcagccgc 9720
 cggcgttagc ccaagagcaa caacagcgc当地 aaggctaccg ctc当地ggcgc gggcacaaga 9780
 acgccatagt tgcttgctt caagactgtg gggcaacat ctccctcgcc cggcgttcc 9840
 ttctctacca tcaacggcgtg gccttccccc gtaacatcc gcatctact cgtcatctt 9900
 acagccccata ctgc当地ccgc ggc当地ggc当地 gccc当地ggca cagc当地ggc当地 cacacagaag 9960
 caaaggc当地 cggatagcaa gactctgaca aagcccaaga aatccacagc ggc当地ggca 10020
 gcaggaggag gagcgtcgc当地 tctggc当地cc aacgaacccg tatc当地ggc当地 cgagctt当地 10080
 aacaggattt tccccc当地tct gtatgtata ttcaacaga gcaaggccca agaacaagag 10140
 ctgaaaataa aaaacaggctc tctgc当地cc ctc当地ccgc当地 gtc当地ctgtt当地 tcacaaaagc 10200
 gaaagatcagc ttccggc当地ac gctggaaagac gccc当地ggctc tctt当地ctaa atactgc当地 10260
 ctgactctt当地 aggacttagtt tccggccctt tctcaattt aagc当地ggc当地 actacgtcat 10320
 ctccagcggc当地 cacacccggc当地 gccagcaccct gtc当地cagc当地 cc当地ttatgag caaggaaattt 10380
 cccacggccctt acatgtggag ttaccagcc当地 caaatggac当地 ttgc当地ggctgg agctgccc当地 10440
 gactactcaa cccgaataaa ctatcgacg gccc当地ggcc当地 acatgtatcc cccggc当地aac 10500
 ggaatccgc当地 cccaccgaaa cc当地attctc当地 ttggaaacagg cggctattac caccacacctt 10560
 cgtatataacc ttaatccccg tagttggccc当地 gtc当地ccctgg当地 tgc当地ccaggaa aagtc当地ccgtt 10620

-continued

cccaccactg tggtaacttcc cagagacgcc caggccgaag ttcatgtac taactcaggg 10680
 ggcagcttg cggggggctt tcgtcacagg gtgcggtegc cggggcaggg tataactcac 10740
 ctgacaatca gagggcggagg tattcagctc aacgacgagt cggtgagctc ctgcgttgt 10800
 ctccgtccgg acgggacatt tcagatcgcc ggcccggcc gtccttcatt cacgctcgt 10860
 caggcaatcc taactctgca gacctcgcc tctgagccgc gctctggagg cattggact 10920
 ctgcaattta ttgaggagtt tgcgtccatcg gtctacttta acccccttc gggacccccc 10980
 ggccactatc cggatcaatt tattcctaacc tttgacgccc taaaggactc ggccggacggc 11040
 tacgactgaa tgtaagtgg agaggcagag caactgcgc tgaaacacct ggtccactgt 11100
 cgccggccaca agtgcgttgc ccgcgcactcc ggtgagttt gctacttga attgcccag 11160
 gatcatatcg agggccggc gcacggcgtc cggcttaccg cccaggaga gcttgcgcgt 11220
 agcctgattc gggagtttac ccagcgcggcc ctgctagttg agcgggacag gggaccctgt 11280
 gttctcaactg tgatttgcaa ctgtcgatct tattccctt aactaataaa aaaaataat 11340
 aaagcatcac ttacttaaaa tcagtttagca aatttctgtc cagtttattc agcagcacct 11400
 cttgccttc ctcccagctc tggatttgc gcttccttc ggctgcaaac tttctccaca 11460
 atctaaatgg aatgtcagtt tcctccgtt cctgtccatc cgcacccact atcttcatgt 11520
 tgttgcagat gaagcgcgc aagaccgtctg aagatacctt caaccccggt tatccatatg 11580
 acacggaaac cggtcctcca actgtgcctt ttcttactcc tccctttgtt tcccccaatg 11640
 ggtttcaaga gagtccccct ggggtactct ctttgcgcct atccgaacct ctgttacct 11700
 ccaatggcat gtttgcgc aaaaatggca acggcctctc tctggacgag gccggcaacc 11760
 ttacctccca aatgttaacc actgtgagcc cacctctcaa aaaaaccaag tcaaacataa 11820
 acctggaaat atctgcaccc ctcacagttt cttcagaagg ccttaactgtg gcttgcgcgg 11880
 cacctctaattt ggtcgccggc aacacactca ccatgcaatc acaggccccg ctaaccgtgc 11940
 acgactccaa acttagcatt gccacccaaag gacccctcac agtgcagaa ggaaagctag 12000
 ccctgcaaac atcaggcccc ctcaccacca ccgatagcag tacccttact atcactgcct 12060
 cacccctct aactactgcc actgttagct tggcatttgc ctggaaagag cccattata 12120
 cacaaaaatgg aaaacttagga cttaagtacg gggctctttt gcatgttaca gacgaccta 12180
 acactttgac cgttagactt ggtccaggtg tgactattaa taataacttcc ttgcaaaacta 12240
 aagttactgg agccttgggt tttgattcac aaggcaatat gcaacttaat gtgcaggag 12300
 gactaaggat tgatttcaaa aacagacgcc ttatacttgc tgtagttat ccgtttgtat 12360
 ctcaaaacca actaaatcta agactaggac agggccctct ttttataaac tcagccacca 12420
 acttggatataactacaac aaaggccctt acttgcgttac agcttcaaac aattccaaaa 12480
 agcttggatgt taacctaagc actgcacagg ggttgcgtt tgacgtaca gccatagcca 12540
 ttaatgcagg agatggcctt gaatttgggtt cacctaactgc accaaaacaca aatccctca 12600
 aaacaaaaat tggccatggc cttagatttgc attcaaacaa ggctatgggtt cctaaactag 12660
 gaaactggcct tagtttgc acgcacagggtt ccattacagt aggaaacaaa aataatgata 12720
 agctaactttt gtggaccaca ccagctccat ctccttaactg tagactaaat gcagagaaag 12780
 atgcgtttactt cactttggc ttaacaaaat gtggcgttca aataacttgc acagtttgc 12840
 ttttggctgt taaaggcgtt ttgggttcaaa tatctggac agttcaaat gctcatctt 12900
 ttataagatt tgacgaaaat ggagtgcata taaacaattt cttccctggac ccagaatatt 12960
 ggaacttttag aatggagat cttactgaag gcacagccata tacaacgcgtt gttggattt 13020

-continued

tgccctaacct atcagcttat cccaaatctc acggtaaaac tgccaaaagt aacattgtca 13080
 gtcaagttt cttaaacgga gacaaaacta aacctgtaa actaaccatt acactaaacg 13140
 gtacacagga aacaggagac acaactccaa gtgcataactc tatgtcattt tcatggact 13200
 ggtctggcca caactacatt aatgaaatat ttgccacatc ctcttacact ttttcataca 13260
 ttgcccaga ataaagaatc gtttgttta tgtttcaacg tgtttatttt tcaattgcag 13320
 aaaatttcaa gtcattttc attcagtagt atagccccac caccacatag cttatacaga 13380
 tcaccgtacc ttaatcaaac tcacagaacc ctgttattca acctgcacc tccctccaa 13440
 cacacagagt acacagtccct ttctccccgg ctggccttaa aaagcatcat atcatggta 13500
 acagacatata tcttaggtgt tatattccac acggtttccct gtcgagccaa acgctcatca 13560
 gtgatattaa taaactcccc gggcagctca cttaaagtca tgtcgtgtc cagctgtga 13620
 gccacaggct gctgtccaaac ttgcgggtgc ttaacgggctg gccaaggaga agtccacgccc 13680
 tacatggggg tagagtctata atcgtgcata aggataggggc ggtggtgctg cagcagcgc 13740
 cgaataaaact gctgcggccg ccgcgtccgtc ctgcaggaat acaacatggc agtggctcc 13800
 tcagcgtatcg ttgcaccgc ccgcagcata aggcgccttg tccctccggc acgcagcgc 13860
 acoctgtatct cacttaaatac agcacagtaa ctgcagcaca gcaccacaat attgttcaaa 13920
 atcccacagt gcaaggcgct gtatccaaag ctcatggggc ggaccacaga acccacgtgg 13980
 ccatcataacc acaagcgcag gtagattaag tggcgaccccc tcataaacac gctggacata 14040
 aacattacct cttttggcat gttgttaatc accacccctcc ggtaccatataa aacccctgt 14100
 ttaaacatgg cgccatccac caccatccta aaccagctgg cccaaacctg cccgcggct 14160
 atacactgcg gggAACGGG actggaaaca tgacagtggc gagcccgagg ctcgtaaacca 14220
 tggatcatca tgctcgatc gatataatg ttggcacaac acaggcacac gtgcatacac 14280
 ttccctcgatc ttacaagctc ctcccggtt agaaccatata cccaggaaac aacccatcc 14340
 tgaatcagcg taaatccccactc actgcaggga agacctcgca cgtaactcac gttgtgcatt 14400
 gtcaaaagtgt tacattcggtt cagcagcgga tgatccctca gtatggtagc gccccggct 14460
 gtctcaaaag gaggttagacg atccctactg tacggagtgc gcccggacaa ccggatcgt 14520
 gttggctgta gtgtcatgcc aaatggaaac ccggacgttag tcataattttcc tgaagcaaaa 14580
 ccagggtcggtt cgcgtacaaa cagatctcgcc tccctccgtt cgcgtttag atcgtctgt 14640
 gtagtagttt tagatatacc actctctcaaa agcatccagg ccgcggccctgg cttccgggttc 14700
 tatgtaaact ctttcattgcg ccgcgtccct gataacatcc accacccgcg aataagccac 14760
 acccagccaa cttttttttt tattccaaaa gattatccaa aacccatccaa tgaagatcta 14880
 ttaagtgaac ggcgtccctt ccggggcgtt ggtccaaactc tacagccaaa gaacagataa 14940
 tggcattttt aagatgttgc acaatggctt cccaaaggca aacccgcctc acgtccaaagt 15000
 ggacgtaaag gctaaaccct tcagggtgaa tccctcttat aaacattcca gcacccatca 15060
 ccatgccccaa ataattctca tctcgccacc ttctcaatata atctctaaacg aaatccggaa 15120
 tattaaagtcc ggcattgtt aaaaatctgtt ccagagcgcc ctccaccccttc agcctcaagc 15180
 agcgaatcat gattgcaaaa attcaggttc ctcacagacc tgtataagat tccaaagcgg 15240
 aacattaaca aaaaatccgc gatcccgtag gtccttcgc agggccagct gaacataatc 15300
 gtgcagggtct gcacggacca ggcggccac ttccccggca ggaaccatga caaaagaacc 15360

-continued

cacactgatt atgacacgc a tactcgagc tatgctaacc agcgtagccc cgatgtaa 15420
 ttgttgcata ggcggcgata taaaatgca a ggtgctgtc aaaaaatcag gcaaagcctc 15480
 ggcggggggaa gaaagcacat cgtagtcatg ctcatgcaga taaaggcagg taagtcgg 15540
 aaccaccaca gaaaaagaca ccattttct ctcaaacatg tctgcgggtt tctgcataaa 15600
 cacaaaataa aataacaaaa aaacatttaa acattagaag cctgtttac aacaggaaaa 15660
 acaaccctta taagcataag acggactacg gccatgcggc cgtgacccgt aaaaaactgg 15720
 tcaccgtat taaaagcac caccgacago tcctcggtca tgtccggagt cataatgtaa 15780
 gactcgtaa acacatcagg ttgattcaca tcggtcagtg ctaaaaagcg accgaaatag 15840
 cccggggggaa tacatacccc caggcgtaga gacaacatta cagccccat aggaggtata 15900
 acaaaattaa taggagagaa aaacacataa acacctgaaa aaccctccgt cctaggcaaa 15960
 atagcacccct cccgctccag aacaacatac agcgcttcca cagcggcagc cataacagtc 16020
 agccttacca gtaaaaaaga aaaccttatta aaaaaacacc actcgacacg gcaccagctc 16080
 aatcagtcac agttaaaaaa agggccaagt gcagagcag tatataatagg actaaaaaat 16140
 gacgtaacgg tttaaagtcca caaaaaacac ccagaaaaacc gcacgcgaac ctacgcccag 16200
 aacgaaagc caaaaaaccc acaacttcct caaatcgta cttccgttt cccacgttac 16260
 gtaacttccc attttaaagaa aactacaatt cccaaacatc acaagttact ccgccttaaa 16320
 acctacgtca acgtttttta aatgtcgacg cgtacaaaac gtcaaaaggg cgacacaaaa 16380
 tttttctaa atgcataata aatactgata acatcttata gtttgtat tattttgtat 16440
 taticgttgc atgtataatt ttgatataaa aactgattt tccctttatt ttttcgaga 16500
 tttttttct taattcttt taacaaacta gaaatattgt atatacaaaa aatcataaat 16560
 aatagatgaa tagtttaatt atagggttca atcaatcgaa aaagcaacgt atcttattta 16620
 aagtgcgttg ctttttctc atttataagg ttaaataatt ctcatatatc aagcaaagt 16680
 acaggcgcgcc ttaaatattc tgacaaatgc tctttccctaa aactcccccc ataaaaaaac 16740
 ccggcgaagc gggttttac gtttttgcg gattaacgt tactcggtt cagaaccgc 16800
 cagggggccc gagcttaaga ctggcgtcg ttttacaaca cagaaagagt ttgttagaaac 16860
 gcaaaaaaggc catccgtcag gggccttctg cttagttga tgccctggcag ttccctactc 16920
 tcgccttcgc ctccctcgct cactgactcg ctgcgtcg tgcgttgcgt gcggcgcagcg 16980
 gtatcagtc actcaaaggc ggtataacgg ttatccacag aatcagggga taacgcagga 17040
 aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg 17100
 gcgttttcc atagggtccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag 17160
 aggtggcgaa aaccgcacagg actataaaga taccaggcgt ttccccctgg aagctccctc 17220
 gtgcgtctc ctgttccgac cctgcgcgtt accggatacc tgcgttgcgtt tctcccttcg 17280
 ggaagcgtgg cgctttctca tagtcacgc tgcgttgcgtt tgcgttgcgtt 17340
 cgctccaagc tgggcgtgt gcacgaaacc cccgttcagc cggacccgtg cgccttatcc 17400
 ggtaactatc gtcttgcgtc caacccgtaa agacacgact tatcgccact ggcagcagcc 17460
 actggtaaca ggatttagcag agcggaggtat gtggcggtg ctacagagtt cttgaagtgg 17520
 tgggctact acggctacac tagaagaaca gtatggta tctgcgtct gctgaagcca 17580
 gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc 17640
 ggtgggtttt ttgtttgca gcaagcagatt acgcgcagaa aaaaaggatc tcaagaagat 17700
 cctttgtatct tttctacggg gtctgacgtc cagtgaaacg acgcgcgcgt aactcacgtt 17760

-continued

```
aagggatttt ggtcatgagc ttgcggcgta ccgtcaagtgc agcgtaatgc tctgctt 17817
```

```
<210> SEQ ID NO 16
<211> LENGTH: 743
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide
```

```
<400> SEQUENCE: 16
```

tgataatccgt	agatgtacct	ggacatccag	gtgatgcggg	cgccgggttgtt	ggaggcgccgc	60
ggaaaagtccgc	ggacgcgggtt	ccagatgttgc	cgcagcggca	aaaagtgcctc	catggtcggg	120
acgctctggc	cggtcaggcg	cgcgcatacg	ttgacgcgtct	agaccgtgca	aaaggagagc	180
ctgttaagcgg	gcaactttcc	gtggtcttgtt	ggataaatttc	gcaagggttat	catggcggac	240
gaccgggggtt	cgagccccgt	atccggccgt	ccgcgcgtat	ccatgcgggtt	accgcggcg	300
tgtcgaaccc	aggtgtgcga	cgtcagacaa	cgggggagtg	ctccctttgg	cttccttcca	360
ggcgccgggg	ctgctcgct	agcttttttg	gccactggcc	gcgcgcagcg	taagcgggta	420
ggctggaaag	cgaaaaggcatt	aagtggctcg	ctccctgttag	ccggagggtt	attttccaag	480
ggttgagtcg	cgggacccc	ggttcgagtc	tcggaccggc	cggactgcgg	cgaacggggg	540
tttgcctccc	cgtcatgcaa	gaccccgctt	gcaaattctt	ccggaaaacag	ggacgagccc	600
ctttttgtt	tttcccaagat	gcatccgggtt	ctggggcaga	tgcgcccccc	tcctcagcag	660
cgccaagagc	aagagcagcg	gcagacatgc	agggcaccc	cccctccctcc	taccgcgtca	720
ggagggggcga	catccgcggt	tga				743

```
<210> SEQ ID NO 17
<211> LENGTH: 3201
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide
```

```
<400> SEQUENCE: 17
```

cgacgttagt	tttagggcg	agtaacttgt	atgtgttggg	aattgttagt	ttcttaaaat	60
gggaagttac	gtaacgtggg	aaaacggaag	tgacgatttgc	aggaagtgtt	gggttttttg	120
gctttcgttt	ctggggctag	gttcgcgtgc	ggttttctgg	gtgttttttg	tggactttaa	180
ccgttacgtc	attttttagt	cctatatata	ctcgctctgc	acttggccct	tttttacact	240
gtgactgatt	gagctggtgc	cgtgtcgagt	ggtgtttttt	taataggttt	tcttttttac	300
tggtaaggct	gactgttatg	gctggcgctg	tggaaagcgct	gtatgtgtt	ctggagcggg	360
agggtgtat	tttgcctagg	caggagggtt	tttcaggtgt	ttatgtgtt	ttctctctta	420
ttaattttgt	tataccctt	atgggggctg	taatgttgc	tctacgcctg	cgggtatgt	480
ttccccccgg	ctatttcggt	cgcttttag	cactgaccga	tgtgaatcaa	cctgtatgt	540
ttaccggagtc	ttacattatg	actcoggaca	tgaccggaga	gtgtcggttgc	gtgcgtttta	600
atcacgggtga	ccagttttt	tacggtcaag	ccggcatggc	cgtagccgt	ctttagctta	660
taagggttgt	ttttcctgtt	gtaagacagg	cttctaattgt	ttaaatgttt	ttttgttatt	720
ttattttgcgt	tttatgcaga	aacccgcaga	catgtttgag	agaaaaatgg	tgtcttttc	780
tgtgggtggtt	ccggagctta	cctgccttta	tctgcgttgc	catgactacg	atgtgtttc	840

-continued

tttttgccg gaggcttcgc ctgatTTTTT gaggcggacc ttgcattta tatcgccgc	900
catgcaacaa gtttacatcg gggctacgt ggttagcata gctccgagta tgctgtcat	960
aatcagtgtg ggTTCTTTG tcatggttcc tggcggggaa gtggccgcgc tggtccgtgc	1020
agacctgcac gattatgttc agctggccct gcgaaggAAC ctacgggatc gcggtatTTT	1080
tgttaatgtt ccgCTTTGA atcttataca ggtctgtgag gaacctgaat TTTGCAATC	1140
atgatTCGTG GCTTGGAGGT gaagggtggag ggccgtctgg agcagatTTT tacaatGCC	1200
ggacttaata ttCGGGATTt GCTTAGAGAT atattgagaa ggtggcggaga tgagaattat	1260
ttgggcattgg ttGAAGGTGC tggaatgttt atagaggaga ttCACCCtGA agggtttagc	1320
ctttaCGTCC acTTGGACGT gaggGCCGTT tgccTTTGG aAGCCATTGT gcaacatTTT	1380
acAAATGCCA ttatCTGTTC ttTGCTGTa gagTTTGACC acGCCACCGG aggggagcgc	1440
gttcaCTTAAG tagatCTTCa ttttgaggTT ttggataATC ttttGGAATA aaaaaaaaaa	1500
catggTTCTT CCAGCTCTTC ccgCTCCtCC cgtgtgtGAC tcgcagaACG aatgtgtAGG	1560
ttggCTGGGT gtggCTTATT ctgcggTGT ggatgttATC agggcAGCgg cgcatGAAGG	1620
agtttACATA gaACCCGAAG ccAGGGGGCG cctggatGCT ttgagAGGT ggataACTA	1680
caactactac acAGAGCGAT ctaAGCGGCG agACCGGAGA CGCAGATCTG tttgtcacgc	1740
ccgcacCTGG tttgCTTCA ggAAATATGA ctacgtCCGG cgTTCCATTt ggcatGACAC	1800
tacgaccaAC acgatCTCGG ttgtCTCGGc GcactCCGta cagtagggat cgtctacCTC	1860
ctttgagAC agAAACCCGC gctaccatac tggaggatCA tccgCTGCTG cccGAATGTA	1920
acactttGAC aatgcacaAC gtgagttACG tgcgaggTCT tccCTGcAGT gtgggATTtA	1980
cgcgtattCA ggaatgggtt gttccCTGGG atatggTTt aacgcgggAG gagCTTGTAA	2040
tcctgaggAA gtgtatGcAc gtgtgcCTGT gttgtGCCAA cattgatATC atgacGAGCA	2100
tgatgatCCA tggTTACGAG tcctgggCTC tccactGTCA ttgttCCAGT cccggTTccc	2160
tgcagtgtat agccgggggg caggTTTGG ccagCTGGT taggatGGT gtggatGGCG	2220
ccatgttAA tcaGAGGTTT atatggTACc gggaggTGGT gaattacaAC atGCCAAAAG	2280
aggtatGTT tatgtCCAGC gtgttatGA ggggtcGCCA cttaatCTAC ctgcgttGT	2340
ggtatGATGG ccacgtGGGT tctgtggTCC ccGCCATGAG ctTTGGATAc agcgcCTTGC	2400
actgtgggAT tttGAACAAT attgtggTGC tgcgtcGAG ttactGTGCT gatTTAAGTG	2460
agatcaggGT gCGCTGCTGT GCCCGGAGGA caaggcgcCT tATGCTGCGG GCGGTGCGAA	2520
tcatcgtGA ggAGACCACT GccatGTTGTt attcCTGcAG gacggAGCGG CGGCGGcAGC	2580
agtttattCG cgcgtGCTG cAGCACCAcC GccCTATCCT gatgcacGAT tATGACTCTA	2640
cccccatGTA GGCgtGACT tctcttCGC cGCCGTTAA GCAACCGCAA GTGGACAGC	2700
agecTGTGGC tcAGCAGCTG gacAGCGACa tGAACtTAAG tgAGCTGCC GGGGAGTTA	2760
ttaatatac tGATGAGCtG ttggotcGAC aggAAACCGT gtggAAATA aCACCTAAGA	2820
atatgtCTGT tacCCATGAT atGATGCTT ttaaggCCAG CGGGGGAGAA aggACTGTGt	2880
actCTGTGTG ttgggAGGGA ggtggcAGGT tGAATACTAG ggttCTGTGA gtttGATTA	2940
ggtaCggGTG tctgtATAAG ctatgtGGTg GTGGGGCTAT actactGAAT gaaaaATGAC	3000
ttgaaatTTT ctgcaattGA AAAATAAACA CGTTGAAACA taacacAAAC gattCTTtAT	3060
tcttggGCAAt tGtATGAAAGA AGTGTAGAG gatgtggcAA atatTCATT aatgtAGTTG	3120
tggccAGACC AGTCCCATGA AAATGACATA gagtagcAC ttggagTTGT gtctCTGTt	3180
tcctgtgtac CGTTAGTGT a	3201

-continued

```

<210> SEQ ID NO 18
<211> LENGTH: 5336
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    polynucleotide

<400> SEQUENCE: 18

caccctgtga cgaaagccgc ccgcaagctg cgccccctgag ttagtcatct gaacttcggc      60
ctgggcgtct ctgggaagta ccacagtggt gggagcggga ctttcctggt acaccaggc      120
agcggggcaa ctacggggat taaggttatt acgaggtgtg gtggtaatag ccgcctgttc      180
caagagaatt cggtttcggt gggcgccgat tccggttgacc cgggatatac tgggggtcc      240
cgcgctcatg tagtttattc gggttgagta gtcttggca gtcggccgca caagtcccat      300
ttgtggctgg taactccaca tgttagggcgt gggaaatttcc ttgctcataa tggcgctgac      360
gacaggtgct ggccgcgggt gtggccgctg gagatgacgt agttttcgcg cttaaatttgc      420
agaaagggcg cgaaaactagt ccttaagagt cagcgcgcag tatTTactga agagagcctc      480
cgcgcttcc agcgtgcgcgaa gaagctgatc ttccgttttgc tgatacaggc agctgcgggt      540
gaggatgcg agagacctgt tttttatTT cagctcttgc tcttggcccc tgctctgttgc      600
aaatatagca tacagagtgg gaaaaatctt gttctaaacgc tcgggggtcg atacgggttc      660
tttggcgcc agacgcagcg ctccctcc tcgtgtgcg cccgctgtgg atttcttggg      720
cttgcaga gtcttgcata cccgtcgctt ttgcgttgcgt tgccggcgtc ctgttgcgtc      780
cgctgcgcgt gcccgggtg cagtatggc tgttagagatg acggtagttaa tgcaggatgt      840
tacggggaa ggccacgcgc tgatggtaga gaagaaagcg gcggggcgaag gagatgttgc      900
ccccacagtc ttgcaagcaa gcaactatgg cgttcttgc cccgcgcacat gagcggtac      960
cttggcgctg ttgttgcata tggcttaacgc gcccggcgtc ctggacttaa cccgcctgg      1020
ttccagtggt gtcccatcta cgggtggcggc ggcgaacggg cagtcggcggc ggcgcctgag      1080
gagcggagggt tgtagccatgc ctggaaacggg ttgcggatt ctggggcgcc ggcgaggggaa      1140
atgcgaccga ggggtacgggt gtttcgtctg acaccttcc gacctcgaa gcttcctcg      1200
ctaggcttc ccaagtcttc atcatgtcct ccccttcgtc gtccaaacc tccctcgct      1260
gactgtccca gtatccctcc tcgtccgtgg gtggcggcggc cagtcgcagc ttcttttgg      1320
gtgccatccct gggaaagcaag gcccggcggc tgctgtgtat agggctggc cggcgggggg      1380
attgggttga gtcctcgcc ggactggggg tccaagtaaa ccccccgtcc ctttcgtac      1440
agaaactctt ggccggctt gttgtggcgt tgcaattggc caagaatgtg gcccggta      1500
atgacgcagg cggtaagctc cgcatttgc gggcgggatt ggtcttcgtaa gacctaatac      1560
tcgtggcgtt ggtgtccatc aggtacaaat ttgcgtggcgt aagccgcacgt ccacagcccc      1620
ggagttagtt tcaaccccg agccgcggc ttttcgtcag gcgaggacc ctgcagctca      1680
aaggtagccgta taatttgcact ttgcgttaacgc agctgcgaat tgcaaaaccag ggagcgggtc      1740
gggggtgcata gggtgcagcg acagtgcacac tccagtagac cgtcaccgcgt cacgtctcc      1800
attatgtcag agtggtaggc aaggttagttg gctagctgca gaaggttagca gtggcccaa      1860
agcggcggag ggcattcgcg gtacttaatgc ggcacaaatgc cgcttaggaag tgcacagcag      1920
gtggcggggca agattcctga ggcgtctaggataaagttcc taaagttctg caacatgtt      1980
tgactggtga agtctggcag accctgttgc agggtttaa gcagggcgttc gggaaaatg      2040

```

-continued

atgtccgcca ggtgcgcggc cacggagcgc tcgttgaagg ccgtccatag gtccttcaag	2100
tttgctta cagttctg cagtccttg aggttgcact cctccaagca ctgctgccaa	2160
acgccccatgg ccgtctgccca ggttagcat agaaataagt aaacgcagtc gcggacgtag	2220
tcgcggcgcc cctcgccctt gagcgtggaa tgaagcactt ttgcccagc gcggtttcg	2280
tgcaaaaattc caaggttagga gaccagggtt cagagctcca cggtggagat cttgcaggg	2340
tggcgtacgt agccctgtcg aaaggtagt tgcaatgtttt cctctagctt gcgctgcata	2400
tccgggttagt caaagaaccg ctgcgtgcac tcaagctcca cggttaacgag cactgcggcc	2460
atcattagtt tgccgtgcctc ctccaagtcg gcaggctcgcc gcgtttgaag ccagcgcgt	2520
agctgctcgat cgccaaactgc gggtagggccc tccctgtttt gtttttgcattt atttgcattt	2580
ctctccagggg gctgcgcacgg gcgcacgtac agtcactca tgactgtgtt catgacccgg	2640
gggggttaggt taagtgcggg gttaggcaag tgggtgaccc cgtatgtcg ttttagtacg	2700
gttagggcgcc cggtgtcacc ctcaggttcc accaacactc cagagtactt ttcattttcg	2760
ctgttttccctt gttgcagagc gtttgcggcg cgcttctcgat cgcttccaaag accctcaaa	2820
atttttggca cttcggttgcgat cgaggcgata tcaggtatgtt cagcgcocctg ccgcaaggcc	2880
agctgcttgtt cccgtcggtt gcccgttggca cggcaggata ggggtatctt gcaatgtttgg	2940
aaaaagatgtt gatagggtggc aagcacctct ggacacggca atacggggta gaatgttgggg	3000
cgcgggttgg gctcgcatgt gcccgtttt tggcgtttgg ggggtacgat cggtgagaat	3060
agggtggcgat cgtaggcaag gctgacatcc gctatggcgat gggccacatc gctgatgtct	3120
tgcaacgcgtt cgcagataat ggccgtactgg cgtatgtcgat gttcaacag cactgcgtct	3180
cccacatcta ggtatcgcc atgccttcc tccccccggc cgcgttgc tctgttgc	3240
tctgcgttgtt cctggtcttg ctttttatec tctgttggta ctgagcggtt ctgtcgat	3300
tcgcttacaa aacctgggtc ctgctcgata atcacttcc tccctctcaag cgggggtggcc	3360
tcgacggggaa aggtggtagg cgcgttggcg gcatcggttgg aggccgttgtt ggcgaactca	3420
gagggggggcc ttaggtgtc ctttttctcg actgacttca tgcgtttttt ctgcctatag	3480
gagaaggaaa tggccagtcg ggaagaggag cagcgcgaaa ccaccccgaa ggcggacgc	3540
gggtcgccgcgac gacgtcccccc aaccatggag gacgtgtcgat ccccgcccccc gtcgeccgg	3600
cctccccccggg cgcggccaaa aaagcggtatc aggcggcgatc tgcgtccga ggacgaggaa	3660
gactcatcac aagacgcgtt ggtccgcgc acacccagcc cgcggccatc gacccgtcg	3720
cgccgatggccatcc ccattgcgc caagaagaaa aagaagcgcc cttctccaa gcccggcg	3780
ccgcacatcac cagaggtaat cgtggacago gaggaaagaaa gagaagatgtt ggcgtacaa	3840
atgggtgggtt tcagcaaccc accgggtgtca atcaagcatg gcaaaaggagg taagcgacaca	3900
gtgcggcgcc tgaatgaaga cgacccatgt ggcgttggta tggcggacgc agaggaagag	3960
gaagagggca gcaagcgaa aagtggaaatt acgggtgtatg acccgcttgcgatc tggccgtat	4020
gtgtctgtcgat gggagaaggcatggaggat ggcgcgcgc tgatggacaa gtaccacgt	4080
gataacgtatc taaaggcgaa cttcaacta ctgcgttgcacc aagtggaaatc tctggcgcc	4140
gtatgcacaa cctggctgaa cgaggaggac cgcgggttgc agctgacccctt caccacgt	4200
aagacccatgt tgacgtatgtt gggcgatcc ctgcggcgatc acctgcgttc gtttgcagat	4260
gtgacccatcaca agcatcacca gcccacgggc tgcgtgttgcgatc ggctgcacccg ctgcgtcg	4320
atcgaaggcgac gacgttacgt tctacacggaa agcattatgt taaataagga gacgtgttgc	4380

-continued

```

gaaatggatg tgacgagcga aaacgggcag cgcgcgctga aggagcagtc tagcaaggcc 4440
agaatcgtga agaaccggtg gggccgaaat gtgggtgcaga tctccaacac cgacgcaagg 4500
tgctgcgtgc acgacgcccgc ctgtccggcc aatcagtttt ccggcaagtc ttgcggcatg 4560
ttcttctctg aaggcgcaaa ggctcagggtg gcttttaagc agatcaaggc ttttatgcag 4620
gcccgtgtatc ctaacgccccca gaccgggcac ggtcacctt tgcgtgcact acgggtgcgag 4680
tgcaactcaa agcctggca cgcgcctttt ttgggaaaggc agctaccaaa gttgactccg 4740
ttcgccctga gcaacgcccga ggacctggac gcggatctga tctccgacaa gagcgtgtg 4800
gccagcgtgc accacccggc gctgtatgtt ttccagtgc gcaaccctgt gtatcgcaac 4860
tcgcgcgcgc agggcggagg ccccaactgc gacttcaaga tatcggccgc cgacctgcta 4920
aacgcgttgg tgcgtgtgc cagcctgtgg agtggaaact tcacccgagc gcccggatg 4980
gttgcgtctg agtttaagtg gagcactaaa caccagtatc gcaacgtgtc cctgcccagt 5040
gcccatacgatgc atgcgcggca gaacccctttt gatTTTaaa cggcgcagac ggcaagggtg 5100
ggggtaataata atcaccggc agtgtacaaa taaaaggatt tgcctttatt gaaagtgtct 5160
ctagtagcatt atttttacat gttttcaag tgacaaaaag aagtggcgtc cctaattctgc 5220
gcactgtggc tgcggaaagta gggcgagtgg cgcgtccagga agctgttagag ctgtccctgg 5280
ttgcgacgca ggggtggcgtg tacctggggc ctgttgagca tggagttggg taccgc 5336

```

<210> SEQ ID NO 19
<211> LENGTH: 810
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 19

```

ttagaaaaac tcatacgagca tcaaatttta ttcatatcag gattatcaat 60
accatattttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga ggcagttcca 120
taggatggca agatcctgggt atcggctctgc gattccgact cgtccaaacat caataacaacc 180
tattaatttc ccctcgtcaa aaataagggtt atcaagttagaa aatcaccat gagtgcac 240
tgaatccgggt gagaatggca aaagttttagt cattttttc cagactgtt caacaggcca 300
gccattacgc tcgtcatcaa aatcactcgc atcaacccaa ccgttattca ttctgtattt 360
cgccctgagcg aggcggaaata cgcgcgtcgtt gttaaaaggc caattacaaa caggaatcga 420
gtgcaaccggc cgcaggaaaca ctgcccagcgc atcaacaata ttttccacccgtt aatcaggata 480
ttcttctaat accttggaaacg ctgttttcc ggggatcgcgatgtt accatgcattc 540
atcaggagta cggataaaaat gcttgcgtt cggaaaggcgc ataaattccg tcagccagtt 600
tagtctgacc atctcatctg taacatcatt ggcaacgcata ctttgcattt gtttcagaaaa 660
caactctggc gcatcgccgtc tcccataaa gcgatagattt gtcgcacccgtt attggccgac 720
attatcgcga gcccatttat acccatataa atcagcatcc atgttggaaat ttaatcgccgg 780
cctcgacgtt tcccggtgaa tatggctcat 810

```

<210> SEQ ID NO 20
<211> LENGTH: 674
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

-continued

<400> SEQUENCE: 20

atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt	60
ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtagg	120
cgaaacccga caggactata aagataaccag gcgtttcccc ctggaaagctc cctcggtgc	180
tctcctgttc cgaccctgcc gcttaccgga tacctgtccg ccttctccc ttggaaagc	240
gtggcggtt ctcatagctc acgctgttagg tatctcagg cggtgttaggt cggtcgctcc	300
aagctggct gtgtgacgca acccccgtt cagcccgacc gctgcgcctt atccggtaac	360
tatcgcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactgg	420
aacaggatta gcagagcgag gtatgttaggc ggtgctacag agttcttcaa gtggggcgt	480
aactacggct acactagaag aacagtattt ggtatctcgc ctctgtgaa gccagttacc	540
ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaac ccaccgctgg tagcggttgt	600
ttttttttt gcaaggcagca gattacgcgc agaaaaaaaaa gatctcaaga agatccttg	660
atcttttcta cggg	674

<210> SEQ ID NO 21

<211> LENGTH: 118

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 21

tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa	60
tgtatTTAGA aaaataaca aataggggtc agtgttacaa ccaattaacc aattctga	118

<210> SEQ ID NO 22

<211> LENGTH: 301

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 22

tattctaat gcataataaa tactgataac atcttatagt ttgttattata ttttgtatta	60
tcgttgacat gtataatTTT gatatcaaaa actgattttc ctttttttat ttgcagatt	120
tatTTCTTA attctcttta acaaactaga aatattgtat atacaaaaaa tcataaataa	180
tagatgaata gtttattat aggtgttcat caatcgaaaa agcaacgtat cttttttaaa	240
gtgcgttgct tttttctcat ttataagggtt aaataattct catatatcaa gcaaagtgac	300
a	301

<210> SEQ ID NO 23

<211> LENGTH: 120

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 23

ggcgccctta aatattctga caaatgctct ttccctaaac tccccccata aaaaaacccg	60
ccgaagcggg tttttacgtt atttgcggat taacgattac tcgttatcag aaccgcccag	120

-continued

<210> SEQ_ID NO 24
 <211> LENGTH: 6838
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 24

ctagagtgcga cctgcaggca tgcaagctt ggcataatcat ggtcatacgat gtttccttg	60
tgaaatttgtt atccgcacac aatccacac aacatacggc ccggaaagcat aaagtgtaaa	120
gcctggggtg cctaattgggtt gagctaactc acattaattt cgttgcgcctc actgcccgt	180
ttccagtcgg gaaacctgtt gtgccagctt cattatggaa tcggccaacg cgccccggaga	240
ggcggtttgc gtattgggcg ctcttcgcgt tccctcgatca ctgactcgct gcgcgcgtc	300
gttcggctgc ggcgagcggt atcagctcac tcaaaggcgta taatacggtt atccacagaa	360
tcaggggata acgcaggaaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt	420
aaaaaggccg cggtgcgtgc gttttccat aggctccgc cccctgacga gcatcacaaa	480
aatcgacgct caagtcagag gtggcgaaccc cgcacaggac tataaagata ccaggcggtt	540
ccccctggaa gtccttcgt ggcgtctcct gttccgacc tgcgcgttac cggataacctg	600
tccgccttc tcccttcggg aagcggtggc ctttcgcata gtcacgcgtg taggtatctc	660
agttcggtgt aggtcggtcg ctccaaagctt ggctgtgtgc acgaaccccc cggtcagccc	720
gaccgcgtcg ctttatccgg taactatcgat cttgagtccca acccggtaaag acacgactta	780
tcgcccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct	840
acagagttct tgaagtgggt gcctaactac ggctacacta gaaggacagt atttggatc	900
tgcgcgtcg tgaagccagt taccttcggg aaaagagttt gtagctcttgc atccggcaaa	960
caaaccaccc ctggtagccgg tggttttttt gtttgcggc agcagattac ggcggaaaaaa	1020
aaaggatctc aagaagatcc tttgatcttt tctacggggctt ctgacgcgtca gtggaaacgaa	1080
aactcacgtt aagggtttt ggtcatgtga ttatcaaaaa ggatcttgc cttagatcctt	1140
ttaaattaaa aatgaagttt taaatcaatc taaagtataat atgagtaaac ttggctcgac	1200
agtttagaaaa actcatcgag catcaaatga aactgcaattt tattcatatc aggattatca	1260
ataccatatt ttgaaaaag ccgtttctgtt aatgaaggag aaaactcacc gaggcagttc	1320
cataggatgg caagatccctg gtatcggtctt ggcattccgg ctcgtccaaatc atcaatacaa	1380
cctattaattt tccctcgatc aaaaataagg ttatcaagtg agaaatcacc atgagtgcgt	1440
actgaatccg gtgagaatgg caaaaatgttca tgcatttctt tccagacttgc ttcaacaggc	1500
cagccattac gtcgtcatc aaaatcactc gcatcaacca aaccgttattt cattcgat	1560
tgccgcgtcg cgagacgaaa tacgcgtatcg ctgttaaaag gacaattaca aacaggaatc	1620
gaatgcaccc ggcgcaggaa cactgccago gcatcaacaa tattttcacc tgaatcgag	1680
tattcttcata atacctggaa tgctgttttcccgatcg cagttgtgttttgc taaccatgc	1740
tcatcaggag tacggataaa atgcttgcgt gtcggaaagag gcataaaatc cgtcagccag	1800
tttagtctga ccatctcatc tgtaacatca ttggcaacgc taccttcacc atgtttcaga	1860
aacaactctg ggcgcgtggg ctcccccatac aatcgataga ttgtcgacc tgattgcgg	1920
acattatcgc gagcccatat atacccatataa aatcgatcgat ccatgttgcgatcattatcgc	1980
ggcctagagc aagacgtttc ccgttgcatac tcccttttcaatattat	2040

-continued

tgaaggcattt atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa	2100
aataaaacaaa taggggttcc ggcacacatt cccccaaaag tgccacactga cgtctaagaa	2160
accattatta acatgacatt aacctataaa aataggcgta tcacgaggcc ct当地tctc	2220
gcccgtttcg gtgatgacgg tgaaaacctc tgacacatgc agctcccgga gacggtcaca	2280
gcttgtctgt aagcggatgc cgggagcaga caagccgctc agggcgcgtc agcgggtgtt	2340
ggcgggtgtc ggggctggct taactatcg gcacagacgc agattgtact gagagtgcac	2400
catatcggt gtgaaataacc gcacagatgc gtaaggagaa aataccgcat caggccat	2460
tcgcattca ggctgcgcaa ctgttggaa gggcgatcg tgccggctc tt当地tattta	2520
cggcagctgg cgaaaggggg atgtgctgca aggcgattaa gttggtaac gccagggtt	2580
tcccaagtca c acggtgtaa aacgacggcc agtgaattcg agctcggtac cc当地aacaaa	2640
agcttatcga gcccggccaa gaggcagtat tt当地tgcaca cgaatacactg gtttatttag	2700
ggtatgcac atgaatgggt taaagggtc gggtaaggta tccgggttccg ataggctgg	2760
tggttctgta ttccccggtg ctgttccggg caaagtccac aaactggggg tccgtttagt	2820
tgtttgtgta ctggatctct gggttccacc tcttggagg tttcttcttgc agctccact	2880
ccatctccac ggtgacctgc ccgggtgtt actgggtat gaagctgctg acggcacgt	2940
ccgagaagct ggtgatattt ccgggcacag gctgtttctt gatgagcatc atggcggtg	3000
ggtgttttag gccgaaatccg cccatggccg gagaggggtg aaagtgcgcc cccgtctctg	3060
ggatcttggc ccagatgggt ccttggagg t acacgtccct ctccatccac acgctgcgg	3120
gcacgatttc ctggaggtt tacgtgccgg tccggggggc agtgggtggag ctctgggt	3180
tggtgccat ctggccggc acgttgcacg ccacgggtt caccggctgc gtctcgctct	3240
cgctgggtat gagcatgtt ccctcgaggt acgtggcggt ggtggccggg tt当地ccggct	3300
ggctgtttaga gatcatagtg ttctccaggg cataggtgtt gctgcctgg aggttgg	3360
tcatgccctt cggctgcggg ggcacctggt aactcgcgc ctcgagctcc atcctatgg	3420
tctgttggaa ggcgcgtaca ctggcgccgt tgaccccgga gcccagggtc cagccctgg	3480
ttcggcccat gggcccccggg aaccagttt tcttaggtgtt ggctgtatctc cc当地ccagg	3540
tcttggtaa ctggactccg ccagtgttat ttgtgctcga gaagcggtac aagtactgg	3600
ccaccagccg gttggccagc ttgaaacaggt tctgactggg agcgaagctg gagtggaaagg	3660
gcacctctc aaagttgtat gtaaaactcaa agttgttgcg ct当地tctcgc atcttgcgg	3720
gaaagtactc taggcagaag aagctgcctc tctcggtggg attttctgtt tt当地cggt	3780
tcagcgtcgc gtaaccgtac tgccggcageg taaagacctg cggagggaaag gccggcaggc	3840
atccctcggtt cccgttgcgg acgacgtagg gcagctggta gtc当地tctcc gtaaaactt	3900
ggacgggttgg ggtgagggtt ttggcgatgg tgggtggtaa gtcctgcacc gtgaccttt	3960
tgacttgaat gttgaagatt ttgactctga gggaccgggg tctgaagccc cagtagttgt	4020
tgatgagtct ttgcccagtct cgggggtctt agtggctgtg gaagcggtaa aagtcaaagt	4080
accccccaggg ggtgctgtat ccaaagtgg cgttggcggtt gttccgtcg acggagccgc	4140
ttttgatctc tcggacttgg tgggtgtgt agctggcagc cacccagggtt cgggtggact	4200
tggtgacgac tctgtcccccc atccacgtgg aatcgcaatg ccaatctccc gaggcatgc	4260
ccactccatc ggcacccctgg ttattgtcgc ccaatggcc gccacccccc gcagacattt	4320
tatcagctcc caaacttgag gctgggtggg ctgggatttg cagctgctgg gatccgtgg	4380

US 12,383,587 B2

99

100

-continued

gtccagcttc ggcgtctgac gaggtggaag gcttggagtc ctctteggc cgagectct	4440
ttcttttgg aaagtggtcg tctatccgt ttccggtagg ggccgtctta gcacccttt	4500
caaccaggcc aaaagggtcg agaacccttt tcttggctg aaagactgcc tttccgaggt	4560
ttcccccgaa ggatgtgtcg tcggcgagct tctcctgaaa ctcggegtcc gcgtggtgt	4620
acttgaggta ggggttgtct cccgcctcaa gctgctcgat gtacgagatg tcgtgtctc	4680
gcccgcaccc gtctgcctg ttgacagatc ctctcgatc gagacgttt ccgggtccga	4740
gatagttata accaggcagc acaagaccac gggcttgatc ttgatgtgc tgattgggtt	4800
ttggtttccg tggggccgct tcaaggccca aaaactcgcg aagacattca ccaacttctt	4860
ccaaccaatc tggagggtga tcaacaaaag acatgactac tcgctttatt tactgttctt	4920
tattggcatc gtcaaaatcc ccaaaatctg acaagtttc ctttccccag gggggatcc	4980
catgacaaat ttgacagtga gttacattgt gacagataca tccattttg ccccgattca	5040
aatattcaca ttcatcacat ttgttagaaa tggatcaaa ttgagcatga tagtcacatt	5100
tgcaatcata ctttgaattc caatttgcg gtcgcagagg agcgggatca acagtcacgt	5160
ctgaactgcg aggccgtctcg ggaacaaaatg agagcctggc ccgttctcc agacttttat	5220
agcttagtatt ggtgacgtca cccagtggc gtttttagaga ttctccgcc cctttagttc	5280
ccgcacccattc cctggaaact ttaaaactcgat gagtccacgg cacctgatgc acctttggcc	5340
aagcaaaaaa gtccttgact tcctgcttag taatcttgcc aaaatctggc gggagccgct	5400
tagtcagttc aaatttgaac atgcggctct ccagcggctg ctgggttca aaggtcggt	5460
aattcccatc caccaccaca cacatgtttg tatttggatg tacaatgaca ggggtagaat	5520
caatttgaac agaggattta catttctgtat cgacccgcac ctttggccccc cccaggatgg	5580
ccttggcgga ttcaaccacc ttgttggtca tctttccctc ctccccacaa atgagcattt	5640
tgtccacaca gtcattaaag ggaaagttt catttgcgatc gttcacgcag ccgtaaaagg	5700
gcacagtgtg ggccatggcc tccgcgtatgt tggcttgcg ggtcgtggcg ggtccgtaga	5760
gccagacgggt gttccctttt ttgaaggago gctgacacca gccgttagagg atggatcccg	5820
cgttagggccgg gtcgttagcca ttcatctcaa aaatttgcata gattctgttt tttgaaatgt	5880
cctcggaaac ggagctcccc acgaggtatg ccaccgcgt ttttgcaga ctcataattt	5940
tggtcgcgtt gtcgagcgcg gccttgcgtatgt ggctccgaga gttccggcg gagttgaagg	6000
agaggtatgt ctccgtattt tcctggatcc actgcgttgcg cggatgtatc cccttgcata	6060
cgagccaccc gaccagctcc atgtacctgg ctgaagttt tgatctgtatc accggcgcat	6120
cagaatttggg attctgatcc tctttgttct gtcctgcgt ctgcgcacacg tgcgatcgat	6180
gctgcgcacac caaccgttta cgcgtccgtga gattcaaaaca ggccgttaaa tactgttcca	6240
tatttagtcca cggccactgg agctcaggatgt gggtttggg gagcaagtaa ttggggatgt	6300
agactctatc caccacccatgg ttcccgccatc cggccgcatt tctggatccc gtgaccgcga	6360
accagtttgg caaagtccgc tcgatccccgc ggttttttttctt ctgaatcagt ttttgcgaa	6420
tctgacttcg gaaacgtccc aaaaccatgg atttcacccccc ggtgggttcc acgagcacgt	6480
gcatgtggaa gtagctctct cccttctcaa attgcacaaa gaaaaggccccc tccggggcct	6540
tactcacacg ggcgcattcc gtcagaaatgt cgcgtccgcag cttctccggcc acggcgtgg	6600
gtgcctgcgc aatcagatcc agatccatgt cagaatctgg cggcaactcc catcttct	6660
cggccacccca gttcacaaag ctgtcagaaa tgccggcag atgctcgatc aggtcgatgg	6720
ggacccttaat cacaatctcg taaaaccccg gcatggccgc tgcgatgc agaagttccct	6780

-continued

atactttcta gagaatagga acttcggaat aggaacttct gatctccgg gggatcca	6838
--	------

<210> SEQ ID NO 25
<211> LENGTH: 816
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 25

ttagaaaaac tcatacgagca tcaaattgtaaa ctgcattttt ttcatatcag gattatcaat	60
accatatttt tgaaaaagcc gtttctgtaa tgaaggagaa aactcaccga ggcagttcca	120
taggatggca agatcctgggt atcggtctgc gattccgact cgtccaaacat caataacaacc	180
tattaatttc ccctcgtcaa aaataagggtt atcaagttagt aaatcaccat gagtgacgac	240
tgaatccgggt gagaatggca aaagtttatg cattttttc cagacttgtt caacaggcca	300
gccatttttgc tcgtcatcaa aatcactcgatc atcaacccaa ccgttattca ttcgtgattt	360
cgcctgagcg agacgaaata cgcgatcgct gttaaaagga caattacaaa caggaatcga	420
atgcacccgg cgccaggaaaca ctgcccggc atcaacaata ttttccacccat aatcaggata	480
tttttttcaat accttggatg ctgtttccc agggatcgca gtggtgagta accatgcattc	540
atcaggagta cggataaaat gctttaggtt cggaagaggc ataaattccg tcagccagtt	600
tagtctgacc atctcatctg taacatcatt ggcaacgcta cttttccat gtttcagaaaa	660
caactctggc gcatacggtt tcccatacaa tcgatagatt gtgcacccat attggccgac	720
attatcgcga gccccatattt acccatataa atcagcatcc atgttggat ttaatcgcgg	780
ccttagagcaa gacgtttccc gttgaatatg gctcat	816

<210> SEQ ID NO 26
<211> LENGTH: 673
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 26

tgtgagaaaa aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcggttt	60
tccataggct ccggcccccgtt gacgagcata aaaaaatcg acgctcaagt cagagggtggc	120
gaaacccgac aggactataa agataccagg cgttttcccc tggaaagctcc ctgcgtcgct	180
ctccgtttcc gaccctgccc cttaccggat acctgtccgc ctttccctt tcggaaagcg	240
tggcgcttcc tcatacgctca cgctgttaggt atctcagttc ggtgttaggtc gttcgttcca	300
agctgggtgt tggcacgaa ccccccgttc agcccgaccg ctgcgcctta tccggtaact	360
atcgtcttga gtccaaaccccg gtaagacacg acttatacgcc actggcagca gccactggta	420
acaggattag cagagcgagg tatgttaggcg gtgtacaga gtttttgaag tgggtggctta	480
actacggcta cactagaagg acagtatttg gtatctgcgc totgctgaag ccagttacct	540
tcggaaaaag agttggtagc tcttgcgttcc gcaaacaaac caccgttgtt agcggtggtt	600
tttttttttttcaat cggcggccat attacgcgcgca gaaaaaaaaagg atctcaagaa gatccttga	660
tcttttttcaat ggg	673

<210> SEQ ID NO 27

-continued

```

<211> LENGTH: 143
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 27

aattgttata cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc      60
tgggggtgcct aatgagttag ctaactcaca ttaattgcgt tgcgctact gcccgtttc      120
cagtcggaa acctgtcgty cca                                         143

<210> SEQ ID NO 28
<211> LENGTH: 5121
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 28

ctgaaggccag ttaccttcgg aaaaagagtt ggtagctt gatccggcaa acaaaccacc      60
gctggtagcg gtggttttt tggggcaag cagcagatta cgcgcaaaaa aaaaggatct      120
caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggAACGA aaactcacgt      180
taagggattt tggcatgtt attatcaaaa aggatcttca cctagatcct tttaaattaa      240
aatgaagtt ttaaatcaat ctaaaagtata tatgagtaaa ctgggtctga cagttagaaa      300
aactcatcga gcatcaaattt aaactgcaat ttattcatat caggattatc aataccatat      360
ttttgaaaaa gccgttctg taatgaagga gaaaactcac cgaggcagtt ccataggatg      420
gcaagatcct ggtatcggtc tgcgattccg actcgccaa catcaataca accttattat      480
ttccccctgt caaaaataag gttatcaagt gagaatcac catgagtgc gactgaatcc      540
ggtgagaatg gaaaaatgtt atgcatttct ttccagactt gttcaacagg ccagccatata      600
cgctcgatcat caaaaatcaact cgcatcaacc aaaccgttat tcattcgta ttgcgcctga      660
gcatcgaa atacgcatgc gctgttaaaa ggacaattac aaacaggaat cgaatgcac      720
cgccgcagga acactgccag cgcatcaaca atatttcac ctgaatcagg atattctt      780
aataaccttgcg atgctttttt cccaggatc gcagtgggtga gtaaccatgc atcatcagg      840
gtacggataa aatgcttgat ggtcgaaaga ggcataaattt ccgtcagcc gtttagtctg      900
accatctcat ctgtaaatc attggcaacg ctaccttgc catgtttcag aaacaactct      960
ggcgcatcg gcttccata caatcgatag attgtcgac ctgattgcc gacattatcg      1020
cgagccattataaccata taaatcaggca tccatgttgg aatttaatcg cggccatag      1080
caagacgtt cccgttgaat atgggtcata ctcttccttt ttcaatatta ttgaaggatt      1140
tatcagggtt attgtctcat gagcgatc atatttgaat gtattttagaa aaataaacaa      1200
ataggggttc cgccacatt tccccggaaaa gtgccacccg acgtctaaga aaccattatt      1260
aacatgacat taacctataa aaataggcgatc acacgaggc ctttcgtct cgccgtttc      1320
ggtgatgacg gtgaaaacct ctgacacatg cagctccgg agacggtcac agcttgcgt      1380
taagcggatg cggggagcag acaagccgt cagggcgcgt cagcgggtgt tggcgggtgt      1440
cggggctggc ttaactatgc ggcacatcg cagattgtac tgagagtgc ccatatgcgg      1500
tgtgaaatac cgcacatcg cgtaaaggaga aaataccgc tcaaggcgc ttcgcattc      1560
aggctgcgcactgttggaa agggcgatcg gtgcgggcctt cttcgatatt acgcccgt      1620

```

-continued

gcgaaagggg gatgtgctgc aaggcgatta agttggtaa cgccagggtt ttcccagtca	1680
cgacgttcta aaacgcggc cagtgaaattc gagctcggtt cccgtaaacaa aagcttatcg	1740
agcggccgca agaggcagta ttttactgac acgaatacac ac ggttattga gggtatgcga	1800
catgaatggg tttaagggtt cgggttaagggt atcgggttcc gataggctcg gtggttctgt	1860
atccccgggt gctgtccggg gcaaagtcca caaaactgggg gtcgtttagt ttgtttgtt	1920
actggatctc tgggttccac ctcttggagt tttccttctt gagctccac tccatctcca	1980
cggtgacctg cccgggtctg tactgggtga tgaagctgt gacgggcacg tccgagaagc	2040
tggtgatatt tccgggcaca ggcgttctt tcatgagcat catggggcgtt ggggttttga	2100
gtccgaatcc gccccatggcc ggagaggggt gaaagtgcgc cccctctt gggatcttgg	2160
cccaagatggg tcccttggagg tacacgtccc tctccatcca cacgctgcgg ggcacgattt	2220
cctggaggtt gtacgtgcgg gtcggggggg cagtggttga gtcgttgcgtt ttgggtggcca	2280
tctggccgcac gacgttgcac gccacggcgt tcaccggctg cgtctcgctc tgcgttgcgtt	2340
tgagcatgtt gcccctggagg tacgtggcgg tggtgccggg gtcgtccggc tggctgttga	2400
agatcatagt gttctccagg gcatagggtt tgcgtccctt gagggttgcgtt gtcgtccgtt	2460
tccggctgggg gggcacctgg taactcgccg cctcgagetc catcctattt gtcgtggcga	2520
aggcgctgac actggcgccgg ttgaccccg gageccaggtt ccagccctgg gtcggccca	2580
tggggcccccgg gaaccagttt ttgttaggtt tggcgatctt cccggccagg ttcttgcgtt	2640
actggactcc gccagtgtta tttgtgtca cgaagcggtt caagtaactgg tccaccagcg	2700
ggttggccag cttgaacagg ttctgactgg gagcgaagct ggagtggaa ggcacccctt	2760
caaagggttta ggtaaactca aagttgttgc ccgttctca gatcttgcgtt ggaaagtact	2820
ctaggcagaa gaagctgctc ctctcggtt gatattctgt gttgtcgccg ttcagcgctc	2880
cgttaacccgtt ctgcggcagc gtaaaagaccc gcccggcagg catccctcg	2940
tcccggttcc gacgacgtt ggcacgtt gtcgttcgtt cgtaaaactt tggacgggttgg	3000
agggtgaggtt gttggcgatg gttgtgttgc agtcctgcac cgtgacccctt ttgacttggaa	3060
tgttgaagat ttgtactctg agggacccggg gtctgaagcc ccagtagttt ttgatgagtc	3120
tttgccttgc tccggggctc cagtggttgc ggaagcggtt aaagtcaaaat taccggcagg	3180
gggtgtgttca tccaaagtag gctgtggcgt tgcttcgtt gacggagccg cttttgcgtt	3240
ctcggtactt gttgtgttgc tagctggca gcaatgggtt tccgggtggac ttgggtgacga	3300
ctctgtcccc catccacgtt gaatcgcaat gccaatctcc cgaggcatgt cccactccat	3360
cggcacccctt gttattgttc cccatggcc cggcacccctt cgcagacatt gtatcgatcc	3420
ccaaacttgc ggtctgggtt gctgggattt gcagctgtt ggttccgtt ggtccagttt	3480
cggcgctgtt ggggttggaa ggcttggagt ccttctcggtt ccgagcccttc ttcttttttgc	3540
gaaagggttca gtctatccgc ttccggtagt gggccgttcc agcaccctt tcaaccaggc	3600
caaaagggttca gagaaccctt ttcttggcctt gaaagactgc ctttccgagg tttccccggaa	3660
aggatgtgttca gtcggcgagc ttcttctgaa actggcgccg cgcgttgcgtt tacttgggtt	3720
agggggttgc tcccgctca agctgctgtt tgtaacggat gtcgttgcgtt cgcgcaccc	3780
cgtctgcctt gttgacaggc tctcttcgtt cggagccgtt tccgggtccg agatagttat	3840
aaccaggcag cacaagacca cgggttgcgtt ctgtatgggtt tttgggttgc	3900
gtggggccgc ttcaaggccc aaaaactcgc gaaagccctt accaacttct tccaaccaat	3960

US 12,383,587 B2

107

108

-continued

ctggagggtg atcaacaaaa gacatgacta ctcgcttat ttactgttct ttattggcat	4020
cgtcaaaatc cccaaaatct gacaagttt cctttccca ggggggaatc ccatgacaaa	4080
tttgacagtg agttacattt tgacagatac atccatTTT gccccattc aaatattcac	4140
attcatcaca tttgttagaa atgttgtcaa attgagcatg atagtcacat ttgcaatcat	4200
accttgaatt ccaattgagc ggTCGAGAG gagcgggatc aacagtcacg tctgaactgc	4260
gaggcgtctc gggAACAAAT gagAGCCTGG cccgcttctc cagacttta tagctagtt	4320
tggtgacgTC acccagtggg cgTTTtagag atttctccgc ccTTtagtt cccgccaatt	4380
ccctggaaAC tttAAACTCG tgagtccacg gcacctgatt gacCTTGC caagcaaaaa	4440
agtccttgac ttcctgctta gtaatcttc caaaatctgg cgggagccgc tttagtcagtt	4500
caaatttgaA catcggtcc tccagcggtc gctgggttc aaaggctgtg gaattccat	4560
ccaccaccac acacatgttt gtattggaaAG ttacaatgac aggggtagaa tcaatttgaA	4620
cagaggattt acatttctga tcgacccgca cctttagccg ccccaggatg gccttggcgg	4680
attcaaccac ctgttggcAt atcttccct cctccacca aatgagcatt ttgtccacac	4740
agtcatTAaaA gggAAAGTTT tcattggtc agttcacgca gccgtAAAAG ggcacagtgt	4800
gggcgatggc ctccgcgtat ttggcttgc cggtcgtggc gggTCCGTAG agccagacgg	4860
tgttccctt gttgaaggag cgctgacacc agccgttagag gatggatccc gcgttagggcg	4920
ggtcgtagcc attcatctca aaaatttgcC agattctgtt ttgtAAatg tcctcgggaa	4980
cggagctccc cacgaggtat tccacccgcG ttttgcgtactcataatt ttggtcgcgt	5040
tgtcgagcgc ggccttgatc tggctccgag agttgccggt ggagttgaag gagaggtatc	5100
tctcctgatt ttcctggatc C	5121

<210> SEQ ID NO 29
 <211> LENGTH: 2175
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 29

ttaaagggtt cgggttaaggat atcggttcc gataggctcg gtgggtctgt attccccgt	60
gctgtccggg gcaaaagtcca caaactgggg gtcgtttagt ttgtttgtgt actggatctc	120
tgggttccac ctcttggagt ttcccttctt gagctccac tccatctca cggtgacctg	180
cccggtgtcg tactgggtga tgaagctgtc gacgggcacg tccgagaagc tgggtatatt	240
tccgggcaca ggcgtgttct tcatgagcat catggccgtt ggggtttga gtccgaatcc	300
ccccatggcc ggagagggggt gaaagtgcgc ccccgatctat gggatcttgg cccagatggg	360
tccttggagg tacacgtccc tctccatcca cacgctgcgg ggcacgatatt cctggaggtt	420
gtacgtgcgc gtcgggggggg cagtggttga gctctgggtt ttgggtggcca tctgcccgc	480
gacgttgcac gcaacgcggc tcacggctg cgtctcgctc tgcgtggta tgagcatgtt	540
gcoctcgagg tacgtggcg gttggccgg gttcgccggc tggctgttga agatcatagt	600
gttctccagg gcatagggtg tgctgcccgt gagggttggt gtcatggcgt tcggctgcgg	660
gggcacctgg taactcgcc cctcgagctc catcctattt gtcgtggcga aggccgtac	720
actggcgcgg ttgaccccggtt agcccaggtt ccagccctgg gttcgccccca tggggcccg	780
gaaccagttt ttgttaggtgt tggcgatct cccggccagg ttcttgcgttga actggactcc	840

US 12,383,587 B2

109**110**

-continued

gccagtgta tttgtgctca cgaagcggtta caagtactgg tccaccagcg ggtaggcag	900
cttgaacagg ttctgactgg gagegaagct ggagtgaaag ggacacctct caaagtgtta	960
ggtaaaactca aagttgtgc ccgttctcgat catcttgcgt ggaaagtact cttaggcaga	1020
gaagctgctc ctctcggtgg gatttctgt gtgtcgccgg ttcagegtcg cgtaaccgtta	1080
ctgcggcagc gttaaagaccc gcggagggaa ggccggcagg catccctcggtcccg	1140
gacgacgtggc ggcagctgg agtgcgtgcgat cgtaaacact tggacgggtgg aggtgagggtt	1200
gttggcgatg gtgggtgggg agtgcgtgcac cgtagctctt tgacttgaa tggtgaagat	1260
tttgactctg agggacccggg gtctgaagecc ccagtagttt tgatgagtc ttgcagtc	1320
tccggggctc cagtggtgtt ggaagcggtt aaagtcaag taccccccagg gggtgctgtta	1380
tccaaagtag gcggtggcgt tgcttccgtc gacggagccg ctgttgcgtt ctcggactgt	1440
gtgggtgttgg tagctggca gcacccagggt tcgggtggac ttggtgacga ctctgtcccc	1500
catccacgtg gaatcgcaat gccaatctcc cgaggcattt cccactccat cgccacccgtt	1560
gttattgtcg cccaatgggc cgccacccctcc cgccagacattt gtatcagctc cccaaacttgc	1620
ggctgggttgg gctgggattt gcagctgtcg ggatccgtcg ggtccagctt cggcgtctgt	1680
cggaggtggaa ggcttggagt ccttttcgtt ccgagcccttc ttcttttttggaaatgtgtc	1740
gtctatccgc ttccggtag gggccgtctt agcaccctct tcaaccaggc caaaagggtt	1800
gagaaccctt ttcttggcctt gaaagactgc cttccgagg ttcccccggaggatgtgtc	1860
gtcggcgagc ttcttctgtt actcgccgtc cgctgggtt tacttgagggt aggggttgtc	1920
tcccgccctca agctgtcgat tgcgtcgat cgcgcgaccc ctgtcgccct	1980
gttgacaggc tctccctcgat cgagaccgtt tccgggtcccg agatagttt aaccaggcag	2040
cacaagacca cgggcttgat cttgtatgtc ctgatgggtt ttgggttcg gtggggccgc	2100
ttcaaggccc aaaaactcgc gaagacccctt accaacttct tccaaccaat ctggagggtt	2160
atcaacaaaa gacat	2175

<210> SEQ ID NO 30
 <211> LENGTH: 3
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 30

cat 3

<210> SEQ ID NO 31
 <211> LENGTH: 6980
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 31

ctagagtcga cctgcaggca tgcaagcttg gctgtatcat ggtcatatgt gtttctgt	60
tgaaattgtt atccgctcac aattccacac aacatacgag ccggaaagcat aaagtgtaaa	120
gcctgggggtt cctaatgggtt gagctaaatcc acatataatggt cgttgcgttc actgcggcgt	180
ttccagtcgg gaaacctgtc gtgcggctt catataatggt tcggccaaacg cgccggggaga	240
ggcggtttgc gtattgggcg ctcttcgtt tccatcgat cttgtcgat cgcgtcgat	300

-continued

gttcggctgc ggcgagcggt atcagctac tcaaaggcgg taatacggtt atccacagaa	360
tcagggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt	420
aaaaaggccg cgttgctggc gttttccat aggctccgcc cccctgacga gcatcacaaa	480
aatcgacgct caagtcaagag gtggcgaaac ccgacaggac tataaagata ccaggcgaaa	540
ccccctggaa gtcctctgt gegetctctt gttccgaccc tgccgcttac cggtatactg	600
tccgccttc tcccttcggg aagegtggcg cttctctata gtcacgctg taggtatctc	660
agttcggtgt aggtcggtcg ctccaagctg ggctgtgtgc acgaaccccc cggtcagccc	720
gaccgctcgcc ctttatccgg taactatcgt cttgagtccaa accccggtaag acacgactta	780
tcgcccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct	840
acagagttct tgaagtggtg gcctaactac ggctacacta gaaggagactt atttggtatac	900
tgcgctctgc tgaagccagt taccttcgga aaaagagttt gtagctcttgc atccggcaaa	960
caaaccaccc ctggtagccgg tggtttttt gtttgcagaagc agcagattac ggcgcgaaaa	1020
aaaggatctc aagaagatcc tttgatcttt tctacgggggt ctgacgatca gtggAACGAA	1080
aactcacgtt aagggatttt ggtcatgtga ttatcaaaaa ggttcttac ctatccctt	1140
ttaaattaaa aatgaagttt taaatcaato taaagtataat atgagtaaac ttggcttgac	1200
agtttagaaaa actcatcgag catcaaatga aactgcaattt tattcatatc aggattatca	1260
ataccatatt ttgaaaaag ccgtttctgt aatgaaggag aaaactcacc gaggcagttc	1320
cataggatgg caagatcccg gtatcggtct gcgttcccgat ctcgtccaaac atcaatacaa	1380
cctattaatt tcccttcgtc aaaaataagg ttatcaagtg agaaatcacc atgagtgcg	1440
actgaatccg gtgagaatgg caaaagtttac tgcatttttcc tccagacttgc ttcaacaggc	1500
cagccattac gtcgtcattc aaaatcactc gcatcaacca aaccgttattt cattcgtat	1560
tgccgcctgag cgagacgaaa tacgtatcgatc ctgttaaaa gacaattaca aacaggaatc	1620
gaatgcaacc ggcgcaggaa cactgccago gcatcaacaa tattttcacc tgaatcgaa	1680
tattttctta atacctggaa tgctgttttcc ccaaggatcg cagtggtgag taaccatgca	1740
tcatcaggag tacggataaa atgcttgcgtg gtggaaagag gcataaaatc cgtcagccag	1800
ttagtgtga ccatctcattc tgtaacatca ttggcaacgc tacctttgcc atgtttgcg	1860
aacaactctg ggcgcattccggtt cttccatc aatcgataga ttgtcgacc tgattgcgg	1920
acattatcgcc gagccattt ataccatataa aatcgatcat ccatgttggaa atttaatcg	1980
ggccttagagc aagacgttcc ccgttgaata tggctcatac tttttttttt tcaatattat	2040
tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgcgtt tatttagaaa	2100
aataaacaataa taggggttcc ggcgcacattt ccccgaaaag tgccacatcg cgtctaa	2160
accattatcataa atcgatcatc aacctataaa aataggcgatc tcacgaggcc ctttcgtctc	2220
gcccgtttcg gtgtatcgatc tgaaaacccctc tgacacatgc agctccggaa gacggatcaca	2280
gttttgtctgt aagcggatgc cgggagcgaa caagcccgatc agggcgccgtc agcgggtgtt	2340
ggccgggtgtc ggggctggct taactatcgatc gcatcgatcg agattgtact gagagtgcac	2400
catatcggtt gtgaaatacc gcacagatgc gtaaggagaa aataccgcatt caggccat	2460
tcgcattca ggctgcgaa ctgttggaa gggcgatcg tgccggccctc ttgcgttata	2520
cgccagctgg cgaaagggggg atgtgctgca aggcgattaa gttggtaac gccagggttt	2580
tcccagtcac gacgttgcataa aacgacggcc agtgaattcg agtcggatcccgat	2640

-continued

ggaaaactaga taagaaaagaa atacgcagag accaaagtcc aactgaaacg aattaaacgg	2700
ttttattgatt aacaagcaat tacagattac gagtcaggta tctggtgcca atggggcgag	2760
gctctgaata cacgccatta gtgtccacag taaagtccac attaacagac ttgtttagt	2820
tggaaagtgtt ctgaatttcg ggattccago gtttgcgtt ttccctctgc agctcccact	2880
cgtatctccac gctgacctgt cccgtggagt actgtgtgat gaaggaagca aactttcccg	2940
cactgaaggt ggtcgaagga ttgcaggtt ccgggggtgtt cttgtatgaga atctgtggag	3000
gagggtgttt aagtccgaat ccacccatga ggggagaggg gtgaaaatgt ccgtccgtgt	3060
gtggaatctt tgccccatgtt ggccccctgaa ggtacacatc tctgtctgtc cagaccatgc	3120
cttggaaagaac gccttgcgtt ttgacatctg cggttagctgc ttgtctgttgc cctctctgg	3180
gttggtaga tacagaacca tactgtccg tagccacggg attgggtgtc ctgatttctt	3240
cttcgtctgtt aatcatgacc ttttcaatgtt ccacattgtt ttctctgttgc ctttgcttcc	3300
caaagatgag aaccccgctc tgaggaaaaa acttttcttc atcgctcttgc tggcttgc	3360
tggccggggcc cggattcacc agagagtcc tcgttgcgtt gttgttgcgtt gtagctccag	3420
tccacgagta ttcaactgtt ttgttatccg cagatgtctt tgataactcgc tgctggcggt	3480
aacagggtcc aggaagccag ttccatgact ggtcccaat gtcactcgtt ccggccttgc	3540
aaaactgaag ctttgcgttgc gtgggttgc cacttggagt gtttgcgttgc ctcaagtaat	3600
acaggtactg tgcgtatgaa ggattcatga gacgggtccag actctggcttgc tgacgttgc	3660
tgctgtggaa aggaacgtcc tcaaaatgtt agctgaaggtt aaagttgtt ccggtaacgca	3720
gcatctgaga agggaaatgtt tccaggcagt aaaatgaaga gcgtccactt gcctgactcc	3780
cgttgcgttccatg ggtgaggtat ccatactgttgc gcaccatgaa gacgttgcgtt gggAACGGCG	3840
ggaggccatcc ttgtatgcgc gagccgagga cgtacggggat ctgggtactcc gagtcagtaa	3900
acacctgaac cgtgcgttgc tggatgttgc caatcgttgc cgttgcgttgc ttctgcgttgc	3960
cctcttttgc ttgtatgttgc aagatgttgc agttgagtctt ctgggttgc aatccccatgtt	4020
tgttgttgc tggatgttgc cagtcacgttgc gtggaaatgtt gcaatgttgc tggatgttgc	4080
caaaatacc ccaagggggtt ctgtatccaa agtagtgc ttttgcgttgc gctccatgttgc	4140
ggctggaaat ttgtttgttgc aggtgggttgc tggatgttgc cagggccatgtt gttgggttgc	4200
tggatgttgc tggatgttgc cccatccatgttgc tggatgttgc atggcaatcc cccatgttgc	4260
tacccactcc gtcggccccc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4320
tctgttgc tggatgttgc cggatgttgc tggatgttgc tggatgttgc tggatgttgc	4380
gtactgttgc tggatgttgc cggatgttgc tggatgttgc tggatgttgc tggatgttgc	4440
ccgccttcc ggttccatgttgc gggatgttgc tggatgttgc tggatgttgc tggatgttgc	4500
ttttccatgttgc acggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4560
ttttccatgttgc acggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4620
gtctgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4680
gtctgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4740
ctccatgttgc ggttccatgttgc aagggttgc tggatgttgc tggatgttgc tggatgttgc	4800
tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4860
accactgttgc ttttgcgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4920
cagccatacc tgatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	4980
accagatgttgc aggttccatgttgc tggatgttgc tggatgttgc tggatgttgc tggatgttgc	5040

-continued

agtttctgat acgcctttt gacgacagaa acgggttgag attctgacac gggaaagcac	5100
tctaaacagt ctttctgtcc gtgagtgaag cagatatttg aattctgatt cattctctcg	5160
cattgtctgc agggaaacag catcagattc atgcccacgt gacgagaaca tttgtttgg	5220
tacctgtctg cgttagttgat cgaagcttcc gcgtctgacg tcgatggctg cgcaactgac	5280
tcgcgcaccc gtttgggctc acttatatct gcgtcactgg gggcgggtct tttcttgct	5340
ccacccttt tgacgtagaa ttcatgtcc acctcaacca cgtgatccct tgcccacccg	5400
aaaaagtctt tgacttcctg cttggtgacc ttcccaaagt catgatccag acgggggtg	5460
agttcaaatt tgaacatccg gtcttgcaac ggctgctggt gttcgaaggt cgttgagttc	5520
ccgtcaatca cggcgcacat gttgggtttg gaggtgacga tcacgggagt cgggtctatac	5580
tggggcaggc acttgcattt ctggccacg cgacacccgc ttccctcggag aatggcttg	5640
gccccactcca cgacccctggc ggtcatcttc ccctccccc accagatcac catcttgtcg	5700
acacagtcgt tgaaggaaa gttcttattt gtccagtttta cgcacccgta gaaggccaca	5760
gtgtgggcta tggccctccgc gatgttgctc ttcccgtag ttgcaggccc aaacagccag	5820
atgggtgtcc tcttgcgaa cttttcgtg gccccatccca gaaagacggaa agccgcata	5880
tggggatcgt acccgtttag ttccaaaatt ttataaatcc gattgtggaa aatgtctcc	5940
acgggctgtc gccccaccag gtatcgcccc gccccgttttag tcaggtctcat aatctttccc	6000
gcattgtcca aggccgcatt gatttggac cgcgcgttgg aggccgcatt gaaggagatg	6060
tatgaggcct ggtccctctg gatecactgc ttctccgagg taatccctt gtccacgagc	6120
cacccgacca gtcctatgtt cctggctgaa gttttgatc tgatcaccgg cgcatcagaa	6180
ttgggattct gattctctt gttctgtcc tgcgtctgc acacgtgcgt cagatgtcg	6240
gccaccaacc gtttacgctc cgtgagatcc aaacagggcgc ttaaatactg ttccatatta	6300
gtccacgcacc actggagctc aggctgggtt ttggggacg agtaattggg gatgtagcac	6360
tcatccacca ctttgcgtcc gcctccggcg ccatttctgg tctttgtac cgccgaaaccag	6420
tttggcaaaag tcggctcgat cccgggtaa attctctgaa tcagttttc gcgaatctga	6480
ctcaggaaac gtccaaaac catggatttcc accccgggtt gttccacgag cacgtgcatt	6540
tggaaagtgc tctctccctt ctcaaattgc acaaagaaaa gggcctccgg ggccttactc	6600
acacggcgcc attccgtcag aaagtgcgc tgcaagttct cggccacggt caggggtgcc	6660
tgcgtcaatca gattcagatc catgtcgaa tctggccggca actcccatc cttctcgcc	6720
acccagttca caaagctgtc agaaatgccc ggcagatgct cgtcaaggcgc gctggggacc	6780
ttaatcacaa tctcgaaaa cccggcatg gcccgtgcgc gttcaaacct cccgcttcaa	6840
aatggagacc ctgcgtgcgc actcgggtt aaatacccg cgtgaccacca tgggtgcga	6900
aaatgtcgca aaacactcac gtgacctcta atacaggacc tccctaaccct tatgacgtaa	6960
ttcacgtcactt gactccacca	6980

<210> SEQ ID NO 32
 <211> LENGTH: 2208
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 polynucleotide

<400> SEQUENCE: 32

ttacagatta cgagtcaggt atctggtgc aatggggcga ggctctgaat acacgcatt 60

agtgtccaca gtaaaagtcca cattaacaga cttgtttagt ttggaagtgt actgaattc	120
gggattccag cgtttgcgt tttccctctg cagctccca ctcgatctcca cgctgacactg	180
tcccgtggag tactgtgtga tgaaggaago aaactttgcc gcactgaagg tggtcgaaagg	240
attcgcaggt accgggggtgt tcttgatgag aatctgtggaa ggagggtgtt taagtccgaa	300
tccacccatg aggggagagg ggtgaaaatg tccgtccgtg tgtggaatct ttgcccagat	360
gggcccccta gaagttacacat ctctgtccctg ccagaccatg cctggaagaa cgccttgtgt	420
gttgacatct gcggttagctg cttgtctgtt gcctctctgg aggttggtag atacagaacc	480
atactgtcc gtagccacgg gattgggtgt cctgatttcc tcttcgtctg taatcatgac	540
cttttcaatg tccacatgg ttttctctga gccttgcctc ccaaagatga gaaccccgct	600
ctgagggaaaa aacttttctt catcgccctt gtggcttgcc atggccgggc ccggattcac	660
cagagagtct ctgccattga ggtggtaactt ggtagctcca gtccacgagt attcactgtt	720
gttggatcc gcaagatgtct ttgatactcg ctgctggggta acagggttc caggaagccaa	780
gttccttagac tggcccgaa tggtaactcg tccggcctga gaaaactgaa gccttgactg	840
cgtgggtgtt ccacttggag tggtaacttctt gctcaagtaa tacaggtaactt ggtcgatgag	900
aggattcatg agacggtcca gacttggct gtgagcgtag ctgctgtgaa aaggaacgtc	960
ctcaaaaatgt tagctgaagg taaagttgtt tccggtaacgc agcatctgag aaggaaagta	1020
ctccaggcag taaaatgaag agcgtccctac tgcctgactc ccgttgcgttca gggtaggtt	1080
tccatactgt ggcaccatga agacgctcg tggaaacggc gggaggcatc cttgtgcgc	1140
cgagccgagg acgtacgggaa gctggtaactc cgagtca gaaacctgaa ccgtgctgg	1200
aagggttattt gcaatcgctg tcgtaccgtt attctgcgtt acctcttgc cttgtatgtt	1260
aaagagctt aagttgagtc tcttgggtcg gaatccccag ttgttgcgttga tgagtcttt	1320
ccagtcacgt ggtgaaaatg ggcagttggaa tctgttgcg tcaaaaatacc cccaagggtt	1380
gctgttagcca aagtagtgtat tggcgatcgaa ggctccctgtat tggctggaaa tttgtttgtat	1440
gagggtgggtt ttgttaggtgg gcaggggcca ggttcgggtt ctgggtgttgc tgactctgtc	1500
gcccateccat gtggaatcgc aatgccaatt tcccggggaa ttacccactc cgtccggcc	1560
ctcgatttgc tctgcatttgc gtgegccact gcctgttagcc atcgatttgc ttcccgacc	1620
agagggggct gctgggtgtt gtccggaggagg ctgggggtca ggtactgagt ctgcgtctcc	1680
agtctgacca aatttcaatc tttttcttgc aggctgttgc cccgccttc cggttcccgaa	1740
ggaggaggctt ggctccacag gagagtgttgc taccggcctc tttttcccg gagccgttt	1800
aacagggttcc tcaaccaggc ccagagggttcc aagaaccctc ttttgcgtt ggaagactgc	1860
tcgtccgagg ttggcccaa aagacgttcc ttctttaagg cgctctgttgc actccgcgtc	1920
ggcggtgggtt tacttgaggt acgggttgc tccgttgcg agctgcgggtt cgtaggctt	1980
gtcgtgttgc agggccggc cgttgcgttgc gttgaccggc tctccctgtt ctagtccgtt	2040
gaagggtccg aggtacttgtt acccagggaa cacaagaccc ctgtgttgc ccttatgccc	2100
ctctgcgggc ttgggtgttgc gtggccagg tttgagcttcc caccactgttcc ttattccctc	2160
agagagatgtt tcctcgagcc aatctggaaatcataaccatcg gcagccat	2208

<210> SEQ ID NO 33

<211> LENGTH: 6893

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

US 12,383,587 B2

119

120

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 33

ctagagtgcga	cctgcaggca	tgcaagcttg	gcgttaatcat	ggtcatacgct	gtttcctgtg	60
tgaaattgtt	atccgcacac	aattccacac	aacatacagag	ccggaaagcat	aaagtgtaaa	120
gcctgggtg	cctaattgagt	gagctaactc	acattaattt	cgttgcgcct	actgcggcct	180
ttccagtcgg	gaaacctgtc	gtgccagctg	cattaatgaa	tcggccaaacg	cgccccggaga	240
ggcggtttgc	gtattggcg	cttccgcgt	tcctcgctca	ctgactcgct	gcgcctcgct	300
gttcggctgc	ggcgagcggt	atcagctcac	tcaaaggcggt	taatacgggtt	atccacagaa	360
tcaggggata	acgcaggaaa	gaacatgtga	gcaaaaaggcc	agcaaaaaggc	caggaaccgt	420
aaaaaggccg	cgttgcggc	gttttccat	aggctccgc	ccccctgacga	gcatcacaaa	480
aatcgacgct	caagtcaagag	gtggcgaaac	ccgacaggac	tataaagata	ccaggcggtt	540
ccccctggaa	gctccctcggt	gcgcctcct	gttccgaccc	tgccgcgttac	cggtacacctg	600
tcogccttcc	tcccttcggg	aagcgtggcg	ctttctcata	gtcacgcgt	taggtatctc	660
agttcggtgt	aggtcgttcg	ctccaagctg	ggctgtgtgc	acgaaccccc	cgttcagccc	720
gaccgctgcg	ccttatccgg	taactatcg	ctttagtcca	acccggtaag	acacgactta	780
tcgcactgg	cagcagccac	tggtaacagg	attagcagag	cgaggtatgt	aggcggtgct	840
acagagttct	tgaagtggtg	gcctaactac	ggttacacta	gaaggacagt	atttggtatac	900
tgcgctctgc	tgaaggcagt	taccttcgga	aaaagagttt	gtagctcttgc	atccggcaaa	960
caaaccaccc	ctggtagcgg	tggttttttt	gtttgcaagc	agcagattac	gcgcagaaaa	1020
aaaggatctc	aagaagatcc	tttgatcttt	tctacgggggt	ctgacgcctca	gtggAACGAA	1080
aactcacgtt	aagggatttt	ggtcatgtga	ttatcaaaaa	ggatcttac	ctagatcctt	1140
ttaaattaaa	aatgaagttt	taatcaatc	taaagtata	atgagtaaac	ttggtctgac	1200
agtttagaaaa	actcatcgag	catcaaata	aactgcaatt	tattcataatc	aggattatca	1260
ataccatatt	tttgaaaaag	ccgtttctgt	aatgaaggag	aaaactcacc	gggcaggatc	1320
cataggatgg	caagatcctg	gtatcggtct	gcgattccga	ctcgcccaac	atcaatacaa	1380
cctattaatt	tcccttcgtc	aaaaataagg	ttatcaagtg	agaaatcacc	atgagtgcac	1440
actgaatccg	gtgagaatgg	caaagtttta	tgcatttctt	tccagacttg	ttcaacaggc	1500
cagccattac	gtctgtcata	aaaatcactc	gcatcaacca	aaccgttatt	cattcgtgat	1560
tgccgcctgag	cgagacgaaa	tacgcgtatcg	ctgttaaaag	gacaattaca	aacaggaatc	1620
gaatgcaacc	ggcgcaggaa	cactgccago	gcatcaacaa	tattttacc	tgaatcagga	1680
tattcttcta	atacctggaa	tgctgttttc	ccagggtatcg	cagtggtag	taaccatgca	1740
tcatcaggag	tacggataaa	atgcttgcgt	gtcgaaagag	gcataaaatc	cgtcagccag	1800
tttagtctga	ccatctcata	tgtaacatca	ttggcaacgc	taccttgc	atgtttcaga	1860
aacaactctg	gcccgcctgg	cttcccatac	aatcgataga	ttgtcgacc	tgattgccc	1920
acattatcgc	gagccatatt	ataccatata	aaatcagcat	ccatgttgg	atthaatcgc	1980
ggcctagagc	aagacgtttc	ccgttgaata	tggctcatac	tcttcctttt	tcaatattat	2040
tgaagcattt	atcagggtta	ttgtctcatg	agcggataca	tatttgaatg	tattttagaaa	2100
aataaaacaaa	taggggttcc	gcccacattt	ccccgaaaag	tgccacactga	cgtctaagaa	2160
accattatta	acatgacatt	aacctataaa	aataggcgta	tcacgaggcc	cttcgtctc	2220

-continued

ggcgcgttcg	gtgatgacgg	tgaaaacctc	tgacacatgc	agctcccgga	gacggtcaca	2280
gcttgtctgt	aaggcgatgc	cggggcaga	caagccccgtc	agggcgcgtc	agcggggtt	2340
ggcggtgtc	ggggctggct	taactatgcg	gcatcagagc	agattgtact	gagagtgcac	2400
catatgcgggt	gtgaaatacc	gcacagatgc	gtaaggagaa	aataccgcat	caggegcac	2460
tcgcccattca	ggctgcgcaa	ctgttggaa	gggcgatcgg	tgccgggcctc	ttcgcttatt	2520
cgcacatgg	cgaaaggggg	atgtgctgc	aggcgattaa	gttgggttaac	gccagggtt	2580
tcccagtcac	gacgttgtaa	aacgacggcc	agtgaattcg	agctcggtac	ccgttagccat	2640
ggaaactaga	taagaaagaa	atacgcagag	accaaagtcc	aactgaaacg	aattaaacgg	2700
tttattgatt	aacaagcaat	tacagattac	gagtcaggta	tctggtgcca	atggggcgag	2760
gtctgtata	cacgcattta	gtgtccacag	taaagtccac	attaacagac	ttgttgtagt	2820
tggaagtgt	ctgaatttcg	ggattccago	gtttgctgtt	ttccttctgc	agctccact	2880
cgcacatccac	gctgacctgt	cccggtggagt	actgtgtgt	gaaggaagca	aactttgccg	2940
cactgaagggt	ggtcgaagga	ttcgcaggta	ccgggggtgtt	cttgcgtgaga	atctgtggag	3000
gagggtgttt	aagtccgaat	ccacccatga	ggggagaggg	gtgaaaatgt	ccgtccgtgt	3060
gtggaatctt	tgcccagatg	ggccccctgaa	ggtacacatc	tctgtctgc	cagaccatgc	3120
ctggaagaac	gccttgcgt	ttgacatctg	cggtagctgc	ttgtctgtt	cctctctgga	3180
ggttggtaga	tacagaacca	tactgctccg	tagccacggg	attgggtgtc	ctgattttct	3240
ttcgcgtgt	aatcatgacc	ttttcaatgt	ccacatttg	tttctctgag	ccttgcttcc	3300
caaagatgag	aaccccgctc	tgaggaaaaa	acttttcttc	atcgcttctg	tggcttgcca	3360
tggccgggcc	cggattcacc	agagagtctc	tgccattgag	gtggtaacttgc	gtagctccag	3420
tccacgaga	ttcactgttgc	ttgttatccg	cagatgtcttgc	tgataactcgc	tgctggcggt	3480
aacagggtcc	aggaagccag	ttccttagact	ggtcccgaat	gtcactcgct	ccggcctgag	3540
aaaactgaag	ccttgactgc	gtgggtgttgc	cacttggagt	gtttgttctg	ctcaagtaat	3600
acaggtactg	gtcgatgaga	ggattcatga	gacggtccag	actctggctg	tgagcgtagc	3660
tgctgtggaa	aggaacgtcc	tcaaaagtgt	agctgaaggt	aaagttgttt	ccggtaacgtca	3720
gcacatcgaga	aggaaagtac	tccaggcagt	aaaatgaaga	gcgtcctact	gcctgactcc	3780
cgttggtag	ggtgaggat	ccatactgtgc	gcaccatgaa	gacgtctgt	ggaaacggcg	3840
ggaggcatcc	ttgatgcgcc	gagccgagga	cgtacggag	ctggtaactcc	gagtcaagta	3900
acacctaaca	cgtgcgtggta	aggattattgg	caatcgctgc	cgtaccgtca	ttctgcgtga	3960
cctcttgcac	ttgaatgtta	aagagcttgc	agttgagtc	cttgggtcg	aatccccagt	4020
tgttggtagat	gagtctttgc	cagtcacgtgc	gtgaaaagtgc	gcagtggaaat	ctgttgaagt	4080
caaaatacc	ccaaggggtg	ctgtagccaa	agtagtgatt	gtcggtcgag	gctcctgatt	4140
ggctggaaat	ttgtttgtag	aggtgggtgt	tgttaggtgg	caggcccag	gttcgggtgc	4200
tggtggtagat	gactctgtcg	cccatccatg	tggaatcgca	atgccaattt	cccgaggaat	4260
tacccactcc	gtcgccgccc	tcgttattgt	ctgcattgg	tgcccccactg	cctgttagcc	4320
tcgtatttagt	tcccagacca	gagggggctg	ctggtggtcg	tccgagaggc	tgggggtcag	4380
gtactgagtc	tgcgctccca	gtctgaccaa	aattcaatct	ttttcttgca	ggctgctggc	4440
ccgccttcc	ggttcccgag	gaggagtctgc	gctccacagg	agagtgtct	accggccctct	4500
ttttcccg	agccgtctta	acaggttctc	caaccaggcc	cagaggttca	agaaccctct	4560
ttttcgctg	gaagactgct	cgtccgaggt	tgcccccaaa	agacgtatct	tctttaaggc	4620

-continued

gtcctgaaa ctccgcgtcg gcgtgggtgt acttgaggta cgggttgtct ccgtgtcga 4680
 gtcgcggtc gtaggctttg tcgtgtcgaa gggccgcggc gtctgcctcg ttgaccggct 4740
 ctcccttgc gagtccgttg aagggtccga ggtacttgta cccaggaagc acaagacccc 4800
 tgctgtcgctc cttatgccgc tctgcgggtt ttgggtgggg tggggccaggt ttgagcttcc 4860
 accactgtct tattccttca gagagagtgt cctcgagcca atctggaaaga taaccatcg 4920
 cagccatacc tgattnaat catttattgt tcaaagatgc agtcatccaa atccacattg 4980
 accagatgc aggcaagtgc acgcgtctggc acctttccca tgatatgtat aatgttagcac 5040
 agtttctgtat acgcctttt gacgacagaa acgggttgag attctgacac gggaaagcac 5100
 tctaaacagt ctttctgtcc gtgagtgaag cagatatttg aattctgatt cattctctcg 5160
 cattgtctgc agggaaacag catcagatcc atgcccacgt gacgagaaca tttgtttgg 5220
 tacctgtctg cgtagttgat cgaagcttcc gcgtgtacgt tcgatggctg cgcaactgac 5280
 tcgcgcaccc gtttgggtc acttatatct gcgtcaactgg gggcggtct tttcttgct 5340
 ccaccccttt tgacgttagaa ttcatgctcc acctcaacca cgtgatccct tgcccacccg 5400
 aaaaagtctt tgacttccctg cttggtgacc ttcccaaagt catgatccag acgggggtg 5460
 agttcaatt tgaacatccg gtcttgcaac ggctgctggt gttcgaaggt cgttgagttc 5520
 cccgtcaatca cggcgcacat gttgggttg gaggtgacca tcacgggagt cgggtctatc 5580
 tggggccgagg acttgcattt ctggtccacg cgcacccctgc ttccctccgag aatggcttg 5640
 gcccactcca egaccccttgc ggtcatcttc ccctccctcc accagatcac catctgtcg 5700
 acacagtctg tgaaggaaa gtttcatttgcgtt ccgttccatccgaa gaaagacgga agccgcatat 5760
 gtgtgggctt tggcctccgc gatgtggc ttccctggtag ttgcaggccc aaacagccag 5820
 atgggtgttcc ttttgcgaa cttttctgt gcccattccca gaaagacgga agccgcatat 5880
 tggggatctg accccgtttag ttccaaattt ttataaatcc gattgtggaa aatgtccctc 5940
 acggggctctt gccccaccag gtatgcgggg gcccgttttag tcaggctcat aatctttcc 6000
 gcattgtcca aggcagccctt gatggggac cgegagttgg aggccgcatt gaaggagatg 6060
 tatgaggccctt ggtccctctg gateactgc ttctccgagg taatccccctt gtccacgagc 6120
 cacccgacca gtcctatgtt cctggctgaa gttttgatcc tgatcaccgg cgcacccgaa 6180
 ttgggattct gattctctt gttctgtcc tcgtgtcgac acacgtcgat cagatgtcg 6240
 gccaccaacc gtttaacgttc cgttagatcc aaacaggcgc tttaataactg ttccatatta 6300
 gtccacgccc actggagctc aggtgggtt ttggggagca agtaattggg gatgttagcac 6360
 tcataccacca ctttgcgttcc gcctccggcg ccatttctgg tctttgtgac cgcacccg 6420
 ttggcaag tggcgtcgat cccgcggtaa attctctgaa tcagtttgc gcaatctga 6480
 ctcaggaaac gtcccaaaac catggatttc accccggtt tttccacgag cacgtgcatt 6540
 tggaaagtgc tctctccctt ctcaaattgc acaaagaaaa gggcctccgg ggccttactc 6600
 acacggcgcc attccgtcag aaagtgcgcg tgcaatccctt cggccacggc caggggtgcc 6660
 tgcgtcaatca gattcagatc catgtcgaa tctggcgcc acctccatcc ctctcgcc 6720
 acccagttca caaagctgtc agaaatgcgcg ggcagatgcgt cgtcaaggc gctggggacc 6780
 ttaatcacaatc tctcgtaaaa ccccgccatg gcccgtgcgc agatcagaag ttccatatact 6840
 ttcttagagaa taggaacttc ggaataggaa cttctgtatcc tccggggat cca 6893

-continued

<211> LENGTH: 179
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 34

```

ccccggcatg gcggtgcgc gttcaaacct cccgttcaa aatggagacc ctgcgtgctc      60
actcgggctt aaataccag cgtgaccaca tggtgtcgca aatgtcgca aaacactcac      120
gtgacctcta atacaggacc tccctaacc tatgacgtaa ttcaacgtcac gactccacc      179
  
```

<210> SEQ ID NO 35
 <211> LENGTH: 6989
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 35

```

ctagagtgcg cctgcaggca tgcaagcttg gcgtaatcat ggtcatagct gtttcctgtg      60
tgaattgtt atccgctcac aattccacac aacatacgg cggaaagcat aaagtgtaaa      120
gcctgggtg cctaatgagt gagctaactc acattaattt cggtgcgc actgcccgt      180
ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgggggaga      240
ggcggttgc gtattggcg ctcttcgctt ccctcgctc ctgactcgct gcgctcgctc      300
gttcggctgc ggcgagcggt atcagctcac tcaaaggcgg taatacggtt atccacagaa      360
tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt      420
aaaaaggccg cggtgcggc gttttccat aggtccgc cccctgacga gcatcacaaa      480
aatcgcgcgta caagtcagag gtggcgaaac ccgcacaggac tataaagata ccaggcggtt      540
ccccctggaa gtccttcgt ggcgttcctt gttccgaccgc tgccgcttac cggataacctg      600
tccgccttc tcccttcggg aagcgtggcg ctttcata gtcacgcgtg taggtatctc      660
agttcgggtt aggtcggtcg ctccaagctg ggctgtgtgc acgaaccccc cggtcagccc      720
gaccgctgcg ctttatccgg taactatcg tttgacttca acccgtaag acacgactta      780
tcgcccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt aggcgggtgt      840
acagagttct tgaagtggtg gcctaactac ggctacacta gaaggacagt atttggtac      900
tgcgctctgc tgaaggcagt taccttcggg aaaagagttt gtagctctt atccggcaaa      960
caaaccaccc ctggtagcgg tggttttttt gtttgcagg agcagattac ggcggaaaa      1020
aaaggatctc aagaagatcc tttgatcttt tctacgggggtt ctgacgcgtca gtggaaacgaa      1080
aactcacgtt aagggatttt ggtcatgtga ttatcaaaaa ggatcttcac cttagatcctt      1140
ttaaattaaa aatgaagttt taaatcaatc taaaatgtat atgagtaaac ttgggtctgac      1200
agtttagaaaa actcatcgag catcaaatga aactgcaatt tattcataatc aggattatca      1260
ataccatatt tttgaaaaag ccgttctgt aatgaaggag aaaactcacc gggcaggatc      1320
cataggatgg caagatcctg gtatcggtct ggcattccga ctcgtccaaatcataacaa      1380
cctattaatt tccctcgctc aaaaataagg ttatcaagtg agaaatcacc atgagtgcacg      1440
actgaatccg gtgagaatgg caaaagttt tgcatttttccatcgacttgc ttcaacaggc      1500
cagccattac gtcgtcatc aaaatcactc gcatcaacca aaccgttatt cattcgtgat      1560
tgcgcctgag cgagacgaaa tacgcgtcg ctgttaaaag gacaattaca aacaggaatc      1620
  
```

-continued

gaatgcaacc ggcgcaggaa cactgccago gcatcaacaa tatttcacc tgaatcagga	1680
tattcttcta atacctggaa tgctgtttc ccagggatcg cagtggtgag taaccatgca	1740
tcatcaggag tacggataaa atgcttgatg gtccgaagag gcataaattc cgtcagccag	1800
tttagtctga ccatctcatac tgtaacatca ttggcaacgc taccttgcc atgttcaga	1860
aacaactctg ggcgcacatggg cttccatac aatcgataga ttgtcgacc tgattgccc	1920
acattatcgc gagccatattt atacccatat aaatcagcat ccatgttggaa atttaatcg	1980
ggcttagaga aagacgttcc cggttgaata tggctcatac ttttccttt tcaatattat	2040
tgaagcattt atcagggtta ttgtctcatg agcgatatac tatttgaatg tatttagaaa	2100
aataaaacaaa taggggttcc ggcacatattt cccggaaaag tgccacactga cgtctaagaa	2160
accattatta acatgacattt aacctataaa aataggcgta tcacgaggcc cttegtctc	2220
gcccgtttcg gtgtatgacgg tgaaaacctc tgacacatgc agctccogga gacggtcaca	2280
gcttgcgtgt aagcggtatgc cggggcaga caagccccgtc agggcgogtc agcggtgtt	2340
ggcggtgtc ggggctggct taactatgcg gcatcagaga agattgtact gagagtgcac	2400
catatgcgggt gtgaaatacc gcacagatgc gtaaggagaa aataccgcattt cagggccat	2460
tcgcccattca ggctgcgcaa ctgttggaa gggcgatcgg tgccggcctc ttctgttatt	2520
cgcacagctgg cgaaaagggggg atgtgctgcg aggacattaa gttgggttac gccagggttt	2580
tcccaagtac gacgttgcgaa aacgacggcc agtgaattcg agtcggatcc cctgtacccat	2640
ggaaaactaga taagaaagaa atacgcagag accaaagttc aactgaaacg aattaaacgg	2700
tttatttgcattt aacaagcaat tacagattac gggtgaggta acgggtgcac atggggcggg	2760
gttcagagta cacgccttct gtatcacatc caaatgttcc acttgttagat ttgttagtagt	2820
tggagggtgtt ctggatctcg ggggtccago gcttgcgttt ttcttctgc agctccatt	2880
caatttccac gctgacactgtt ccgggtctgtt attgcgtat gaaagagttc agctttgact	2940
gggttgaaggt ggtcgaggaa tccgcaggta caggcggtttt ctgtatcagg atctgaggcg	3000
gaggatgtt caggccaaag ccgcctatca gcggagacgg gtggaaagtgc ccgtccgtgt	3060
gaggaatctt ggcccaatgtt ggaccctgcg ggtacacgtc ccgggtctgc cagaccatac	3120
cggttgcggc cccctggctg ttgacagtgc caatttgcgg agccgtgtt tgctgcgtca	3180
agttatctgc cacgataccg tattctctgc tagccacagg gttatgtgtt ttgatttttt	3240
cctcgctgtt gacatgaca tcgctgtat ccgcattgtc tctggcagca ttttttttgc	3300
caaaaaatcg gatcccgtta ctggaaaaaa aacgcttcctc gtcgttttg tggatgttgc	3360
tagcgatgcc aggattagcc aatgaatttc ttccattcag atggatatttgc tcccgacag	3420
tccaggcaaa gttgttattt ttgttttgcg cgggttgcgt tgagacgcgt tggtggcggt	3480
aacagggtcc tggcagccag ttcttgcgtt gatggccat tggatgttgc ccaccttggc	3540
tgaagcccgag agtctgcgtt ttggccgttgc ctccgttgcgtt tggatgttgc gacaagtagt	3600
acaggtactg gtcaatcaga ggattcatca gccggtccaa gctctggctg tggcgtagc	3660
tgtgtggaa aggacacgtcc tcgaaagggtt aagtaacttgc gaaatgttgcg tgggttctca	3720
gcacatctgcgaa agggaaatgtt tccaggcgtt agaaggaggaa gctccacgc gcctgactac	3780
cggttgcgtt gtttaggttgcgtt ccgtactggg gaaatcatgaa cacgtccggcc gggAACGGAG	3840
gcaggcagcc ctgggtggca gagccgagaa cgtacggcag ctggatctcc gatgttgcgtaa	3900
acacctggat ggtgttgcgtt aggttatttgcgatgttgcgtt ggtgccttca ttctgcgtca	3960

US 12,383,587 B2

129

130

-continued

cctccttgac ctggatgtt aagagcttga agctgagtct cttggcccg aatccccagt	4020
tgttgttat gagtcgctgc cagtcacgtg gtgaaaagtgc cagtgaaat ctgttaaatgt	4080
caaaataccc ccagggggtg ctgtagccga agtaggtgtt gtcgttggc gtcctcccg	4140
atgtcccggtt ggagatttgc ttgttagaggt ggttgggtt ggtggcagg gcccagggtt	4200
gggtgttgtt ggtgtactctgt ctgtcgccca gcatgttggaa atcgcaatgc caatcccg	4260
aggaactacc cactccgtcg gcgccttcgtt tattgtctgc cattgggtcg ccacccgtcg	4320
cagccattgtt attaggtccc acaccagagg gcgcgtgtgg aggttctccg agaggtttag	4380
gggtgttgtt ggtgtactctgt tcgtcgccca gcatgttggaa atcgcaatgc caatcccg	4440
gttggccctt cttggccatgtt cccgttaggg agtctggaga acgctgggtt gatggctta	4500
ccggtctctt ctttccagga gccgtcttag cgccttcctt aaccagaccc agaggtttag	4560
gaacccgctt cttggccctgg aagactgctc gcccgggtt gcccccaaaa gacgtatctt	4620
cttgcagacg ctccctgaaac tcggcgctgg cgtgggtata ccgcaggatc ggattgtcac	4680
ccgcctgcag ctgcgttgtcg taggccttgc cgtgcgtcg ggcgcgtcg tccgcgcgt	4740
tgacgggctc ccccttgtcg agtccgttga agggtcccgag gtactttagt ccaggaagca	4800
ccagaccccg gccgtgtcc tgctttgtt ggttggctt gggcttcggg gtcgggtt	4860
tcagcgccca ccactcgcgaa atgcctcgtt agagggttgc ctcgagccaa tctggaaat	4920
aaccatcgcc agccataacctt gatttaaatc atttattgtt caaagatgca gtcatccaaa	4980
tccacatgtt ccagatcgca ggcagtgtcaa ggcgtctggca ctttccat gatatgtga	5040
atgttagcaca gtttctgata cgccttttgc acgacagaaaa cgggttggaa ttctgacacg	5100
ggaaaggact ctaaacagtc ttctgtccg tgagtgaagc agatatttga attctgattt	5160
attctctcgat ttgtctgca gggaaacago atcagattca tgcccacgtt acgagaacat	5220
ttgttttgtt acctgtctgc gtatgttgcg gaagcttccg cgtctgacgt cgatggctgc	5280
gcactactgc cgcgcaccccg tttgggtca cttatatgtt cgtcaactggg ggcgggtctt	5340
ttcttggctc caccctttt gacgtagaat tcatgttcca cctcaaccac gtatccctt	5400
gcccacccggaa aaaagtcttt gacttccgtt ttgggtgaccc tcccaaagtc atgatccaga	5460
ccggcgggtga gttcaaattt gaacatccgg tcttgcacgtt gctgtgggtt ttcgaaggc	5520
gttggatccc cgtcaatcac ggcgcacatg ttgggtgtgg aggtgacat cacggggatc	5580
gggtctatctt gggccgagga cttgcatttgc tggtccacgc gcacccgtt tccctccgaga	5640
atggcattttgg ccgactccac gaccttggcg gtcacatccca cctccatccca ccagatcacc	5700
atcttgcgtt ctcgttgcgtt gaagggaaat ttctcattgg tccagtttac gcacccgtt	5760
aaggggccacag tggggctat ggcctcccgat atggtggctt tcccggtatg tgcaggccaa	5820
aacagccaga tgggtttccctt cttggccaaat ttttctgtgg cccatcccg aaagacggaa	5880
ggcgcatattt ggggatcgta cccgtttatgtt tccaaaattt tataaatcccg attgtggaa	5940
atgtcccttca cgggtgtctg gcccaccagg tagtcgggggg cgggtttatgtt caggtctata	6000
atcttcccgat cttgtccaa ggcacccgtt atttggacc gcgagttggaa ggcgcatttgc	6060
aaggagatgtt atgaggccttgc gtccttccgtt atccactgtt tctccgaggtt aatcccccttgc	6120
tccacgagcc acccgaccatg cttccatgttac ctggctgttgcg ttttgcgtt gatcaccggc	6180
gcacatgttgcg tgggtttccgtt atcttccgtt ttcgttgcgtt ggcgttgcgtt ctcgttgcgtt	6240
agatgtgtcg ccaccaaccgc tttacgttcc gtgagattca aacaggccgt taaatactgt	6300
tccatatttttgc tccacgccccca ctggagctca ggctgggttt tggggagccaa gtaattgggg	6360

-continued

```

atgttagcaactatccaccac cttgttccccg cctccggcgcc catttcttgtt ctgttgcacc 6420
gcgaaccagt ttggcaaaagt cggtcgata ccgcggtaaa ttctctgaat cagttttcg 6480
cgaatctgac tcaggaaaacg tccaaaacc atggattca ccccggttgtt ttccacgagc 6540
acgtgcgtgtt ggaagtagct ctctcccttc tcaaattgc caaagaaaag ggcctccggg 6600
gccttactca cacggcgcca ttccgtcaga aagtgcgcgt gcagcttctc ggccacggc 6660
aggggtgcct gctcaatcag attcagatcc atgtcagaat ctggcgccaa ctcccatcc 6720
ttctcgccca cccagttcac aaagctgtca gaaatgcggg gcagatgcgc gtcaaggtag 6780
ctggggacct taatcacaat ctcgtaaaac cccggcatgg cggctgcgcg ttcaaaccctc 6840
ccgcttcaaa atggagaccc tgcgtgctca ctggggctta aatacccagc gtgaccacat 6900
gggtgtcgcaa aatgtcgaa aacactcagc tgacctctaa tacaggaccc ctctaaccct 6960
atgacgtaat tcacgtcagc actccacca 6989

```

```

<210> SEQ ID NO 36
<211> LENGTH: 2217
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

```

```

<400> SEQUENCE: 36
ttacagatta cgggtgaggta aacgggtgcc aatggggcggtt cagagt acacgccttc 60
tgttataaca gcaaaagtccca cacttgtaga tttgttagtag ttggagggtgt actggatctc 120
gggggttccag cgcttgctgtt tttccttctg cagctcccat tcaatttcca cgctgacctg 180
tccgggtgctg tattgcgtga tgaaaagagtt cagctttgac tgggttgaagg tggtcggagg 240
atccgcaggta acaggcgtgt tcttgatcag gatctgaggc ggaggatgtt tcaggccaaa 300
ggcccccacatc agcggagacg ggtggaaagtt gccgtccgtg tgaggaatct tggccca 360
gggaccctgc aggtacacgt cccgggtctg ccagaccata cgggttaagg cccctggct 420
gttgacagtt ccaattttagt gagccgtgtt ttgctgctgc aagttatctg ccacgataacc 480
gtattccctt gtagccacag ggttagtggtt tttgattttctt tcctcgctgg tgagcatgac 540
atcgctgtaa tccgcattgt ctctggcagc attttgggtt ccaaaaatca ggatcccgtt 600
actggggaaa aaacgcgtcct cgtcgatctt gtgtgttgc atagcgatgc caggattagc 660
caatgaattt cttccattca gatggattttt ggtcccagca gtccaggccaa agttgttatt 720
gttggggcccgcc tggatggccca ttgttattagg cccaccttgg ctgaagccca gagttctgcgt 780
atttggccgtg cctctgttg tttgagtcgg agacaagtag tacaggtact ggtcaatcag 900
aggatttcatac agccggccca agctctggct gtggggcttagt ctgctgtggaa aaggcacgtc 960
ctcgaagggtg taagtaaact ggaagttgtt gccgggttctc agcatctgcg aaggaaagta 1020
ttccaggccatc tagaaggagg agcgtccac ggcctgacta ccgttggta gtgttagta 1080
ggccgtactgg ggaatcatga acacgtccgc cgggaacggc ggcaggccagc cctgggtggc 1140
agagccgaga acgtacggca gctgggtactc cgagtccgtaa aacacctggta tggtgttgtt 1200
gaggttatttgc cgcgttgttctt attctgcgtt acctcccttgc cctggatgtt 1260
gaagagcttgc aagctgagtc tcttggcccg gaatccccag ttgttggta tgagtgcgt 1320
ccagtcacgtt ggtaaaaagt ggcagtggaa tctgttaaag tcaaaaatacc cccagggggt 1380

```

-continued

gctgttagccg aagttaggtgt tgcgttgggt ggctcctccc gatgtcccgatggagattg	1440
ctttagagg tgggttgtgt aggtggccag ggcccaaggat cgggtgtctgg tggatgtac	1500
tctgtcgccc agccatgtgg aatcgcaatg ccaatttccc gaggaactac ccactccgtc	1560
ggcgccctcg ttattgtctg ccattggtgcc gccaccggct gcagccatgtt taggtcc	1620
cacaccagag ggcgtgtcg gagggtctcc gagaggttga gggcttgaa ctgactctga	1680
gtcgccagtc tgacaaaaat tgagtcttt tctggccggc tgggtggccctt tcttgcgcgt	1740
gccccgttagag gagtctggag aacgtgggg tggatgtatcc accgggtctt tctttccagg	1800
agccgtctta gcgccttcc caaccagacc gagaggttgc agaaccggct tcttggctg	1860
gaagactgtc cgcccgaggt tgcccccaaa agacgtatct tcttgcagac gctcctgaaa	1920
ctcgccgtcg cgctgggttat accgcaggta cggattgtca cccgcctgca gctgtggtc	1980
gtaggcccttg tcgtgtcga gggccggctgc gtccggcccg tggacgggct ccccttgc	2040
gagtccgttg aagggtccga ggtacttgc gccaaggaaacc accagacccc ggccgtcg	2100
ctgctttgc tgggtggctt tgggtttcg ggtccaggat ttcagccccc accactcg	2160
aatgcctca gagaggtgt cctcgagcca atctggaaaga taaccatcg cagccat	2217

<210> SEQ ID NO 37
<211> LENGTH: 6983
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 37	
ctagagtgcg cctgcaggca tgcaagcttg gcgttaatcat ggtcatagct gtttctgt	60
tgaaattgtt atccgctcac aattcccacac aacatacgag ccggaaagcat aaagtgtaaa	120
gcctggggtg cctaatacgatc gagtaactc acattaatttgcgttgcgtc actgcggct	180
ttccagtcgg gaaacctgtc gtgcaggctg catatatgaa tcggccaaacg cgccgggaga	240
ggcggtttgc gtattggggcg ctcttcgcgt tcctcgctca ctgactcgatc gcgctcg	300
ttccggctgc ggcgagcggt atcagctcac tcaaaggccgtaatacggtt atccacagaa	360
tcaggggata acgcaggaaaa gaacatgtga gcaaaaggcc agcaaaaggcc caggaaccgt	420
aaaaaggccg cggtgtggc gttttccat aggctccgc cccctgacga gcatcacaaa	480
aatcgacgtc caagtcaagat gttggcggaaac ccgcacaggatataaagatcccggttt	540
ccccctggaa gtccttcgtt gctcttcgtt gttccgaccc tgccgttac cgatcacctg	600
tccgccttc tcccttcggg aagcggtggc ctttcgtatcgtatcgtcgttgcgttgcgtt	660
agttcggtgtt aggtcggtcg ctccaaatgtc gggtgtgtgc acgtaccccgatcccg	720
gaccgcgtcg cttatccgg taactatcgatcgttgcgttca acccggtaaac acacgactta	780
tcggccactgg cagcagccac tggtaacagg attagcagag cgaggtatgtt aggcgggt	840
acagagttt tgaagtgggt gcctaaatcggatc ggctacacta gaaggacgtt atttggatc	900
tgcgtctgc tgaaggccatgt taccttcggaaa aaaaatgttgcgttgcgttgcgttgcgtt	960
caaaccaccc ctggtagccgg tggttttttt gtttgcaggc acgaggatcc ggcggccaaa	1020
aaaggatctc aagaagatcc tttgtatctttt tctacgggggt ctgacgttca gttggaaacggaa	1080
aactcacgtt aagggtttt ggtcatgtga ttatcaaaaaa ggatcttccat ctagatccctt	1140
ttaaattaaa aatgaagttt taaatcaatc taaatgtatat atgagtaaac ttgggtctgac	1200

-continued

agtttagaaaa actcatcgag catcaaatga aactgcaatt tattcatatc aggattatca	1260
ataccatatt ttgaaaaag ccgttctgt aatgaaggag aaaactcacc gaggcagttc	1320
cataggatgg caagatcctg gtatcggtct gcgattccga ctcgtccaac atcaatacaa	1380
cctattaatt tcccctcgta aaaaataagg ttatcaagt agaaatcacc atgagtgacg	1440
actgaatccg gtgagaatgg caaaagtttgcatttccagacttgc ttcaacaggc	1500
cagccattac gtcgtcatc aaaatcactc gcatcaacca aaccgttatt cattcgtgtat	1560
tgcgcctgag cgagacgaaa tacgcgtatcg ctgttaaaag gacaattaca aacaggaatc	1620
gaatgcaacc ggccgcaggaa cactgccagc gcatcaacaa tattttcacc tgaatcgaa	1680
tattcttcta atacctggaa tgctgtttccagggatcg cagtggtgag taaccatgca	1740
tcatcaggag tacggataaa atgcttgatgcgttgcagag gcataaaatccgtcagccag	1800
tttagtctga ccatctcatc tgtaacatca ttggcaacgc taccttgc atgtttcaga	1860
aacaactctg ggcgcacgggg cttccatatac aatcgataga ttgtcgacc tgattgcgg	1920
acattatcgc gagccattt ataccatat aaatcagcat ccatgttggaa atttaatcgc	1980
ggcccttagc aagacgtttcccggttgcata ttggcatac tttcccttttcaatattat	2040
tgaagcattt atcagggtta ttgtotcatg agcggataca tatttgaatg tatttagaaa	2100
aataaaacaaa taggggttcc ggcacattt cccccaaag tgccacactga cgtctaagaa	2160
accattatata acatgacatt aacctataaa aataggcgta tcacgaggcc ctttgtctc	2220
ggcggttcg gtatgcggc tgaaacatc tgacacatgc agctccggaa gacggtcaca	2280
gcttgcgtgt aagcggatgc cgggagcaga caagcccgatc agggcgcgtc agcgggtgtt	2340
ggcggtgtc ggggctggct taactatcgatc gcatcagacg agattgtact gagagtgcac	2400
catatgcgtt gtgaaatacc gcacagatgc gtaaggagaa aataccgcatac caggegcata	2460
tcgcccattca ggctgcgaa ctgttggaa gggcgtatcgatc tgccggccctc ttgcgttata	2520
cggcagctggc gaaaaagggggg atgtgtctgc aggcgattaa gttgggtac gcccagggtt	2580
tcccaactc acgtgttgc aacgcacggc agtgaattcg agctcggtac ccgttagccat	2640
ggaaactaga taagaaagaa atacgcagag accaaagtgc aactgaaacg aattaaacgg	2700
tttatttattt aacaagcaat tacagattac gagtcaggta tctggtgccat atggggccgg	2760
gttcaactata tacacattca gtatcacatc caaattcaac attattagac ttgtatagtc	2820
tggaaatgttgcgttgc gcttgcgtgtt ttccctctgc agctccact	2880
cgatctccac gctgacttgc ccagtagaat actgggtgat gaaagagtgc agcttgcct	2940
tgttgaaggc cgttggagga tccgcaggta caggtgtgtt ttgtatgagg atctgaggag	3000
gcgggtgtt cattccaaac cctccatca gcccggaaagg gtgaaatgtgc cctgtgtgt	3060
gaggaatttt gccccaaatgggtccatca ggtacacatc tctgtctgc caaaccatac	3120
ccggaaatgtat tccctgggtt tgaacccago cggtctgcgc ctgtgttgc gcaactctgg	3180
ggtttgcgttgc cacttgcata taggactccgat ttgttgcgttgc gtttagtagtt ttaatttctt	3240
cttgcgttgcgtt tatcatgact ttgttgcgttgc ccatcgttgc tttccatgtt ccttgcgttgc	3300
caaaaattaa agatccagac aaaggaaaga aacggccttc tccttcttg tggctggcca	3360
tagcaggatcc aggattcatc aagcttattac gtccattgcg agcccaagaa gaagctccag	3420
gccaagcaaa ttgcgttgcgttgc gtttttgcgttgc tcaatgttgcgttgc gttgtcggt	3480
agctgggtcc aggtatgttag tttttccatca ggacagccat gttgtgggtt ccggccacac	3540

US 12,383,587 B2

137

138

-continued

tgaattttag cgtttggta ttctgtccag aaccgttaat agtcttttag agataagtaca 3600
agtattggtc gatgagtggta ttcatttagtc ggcccaggct ttggctgtga gcgtactgc 3660
tatggaaagg tacgttctca aactcgtagc tgaactggaa gttgttaccc gttcttagca 3720
tttgcgacgg gaaatattcc aggccagtaaa aggacgaacg acccaeggcc tggctccat 3780
cattaagcgt cagatacccg tactgaggaa tcatgaaaac gtccgctggg aacggcgga 3840
ggcagccctc gtgagccgac ccgagcacgt acgggagctg atagtcgtag tccgtgaaga 3900
cctggaccgt gctggtaagg ttattggcga tggcttgcac tccattgttgc tccgtaacct 3960
cttgcacgt aatgttgaag agcttgaagt tgagtcgtt agggccgaaat ccccgatgt 4020
tgttgcgtgag tcgctgccag tcacgtggtg agaagtggca gtggaatctg ttgaagtcaa 4080
aataccccca ggggggtgcgt tagccgaagt aggcgttgcatttgaagat cctccagatg 4140
tgctgttggg gatttgcttg tagaggtgat tttgttaggt gggcaggccc caggttcggg 4200
tgctgggtgt gatgactctg tccccccagcc attgggaatc gcaatgcacaa ttccggagg 4260
aactaccaccc tccatccggca cttcggttat tgcgtccac tggtgccca ccacctgaag 4320
ccattgttaag agatccacca cctggggggg ctggggggg ttctccgatt gggttgggg 4380
ctggggactga ctctgtgtcg ccagtcgtac cgaaatttgc tctctttta gcgggctgtg 4440
caccggattt gccaataaccc ggggggggtt ccgggttccgtt aggagactgc tctacaggcc 4500
tcttcttcc agggccgtc ttageccgtt cctcaaccag accaagaggt tcaagaagcc 4560
tcttttggc ctgggaact gctcgcccgaa ggttgcggccaaa aaaaagacgtatcttgc 4620
ggcgctactg gaactccggcg tcggcggtt tttacttgcgt gtaacgggttgc tctccggcct 4680
tgagctgtgt gtcgtggcc ttgtcggtgtt cgagggccgc cgcgtctgttgc gctttggcc 4740
gtccccctt gtcgagtcgtt ttgggggggtt caaggatattt gtaacccggaa agcacaagac 4800
ctcgaggggtt gtcttgcgtt ttttttttttgcgttccggccca ggtttcaagcc 4860
ccaccaccc gcaatttcgtt tcaactaaggat ttttttttttgcgttccggccca ggtttcaagcc 4920
ccggcggccat acctgtttaa aatcatttttgcgttccggccca ggtttcaagcc 4980
ttgaccagat cgccggccgtt gcaaggcttgc ggcacccccc ccatgatatg atgaatgttag 5040
cacagtttgcgtt gatacgcctt ttttgcgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5100
cactcttaac agtctttctgttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5160
tcgcattgttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5220
ttgttacctgttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5280
gactcgccgttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5340
gtccaccccttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5400
cgaaaaaaaggcttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5460
gtgagttcaatccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5520
ttcccgtaatccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5580
atctggggccatccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5640
ttggccgactccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5700
tcgcacacgttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5760
acagttgtggccatccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5820
cagatgggttccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5880
tattggggatccgttccggccca gaaacgggtt gagatttgcgttccggccca ggtttcaagcc 5940

US 12,383,587 B2

139

140

-continued

tccacgggct gctggcccac caggttagtcg ggggcggttt tagtcaggct cataatctt	6000
cccgcatgtt ccaaggcagc cttgatgg gaccgcgagt tggaggccgc attgaaggag	6060
atgtatgagg cctggtcctc ctggatccac tgcttctccg aggtaatccc cttgtccacg	6120
agccacccga ccagctccat gtacctggct gaagtttttgc atctgatcac cgccgcacatca	6180
gaattggat tctgattctc tttgttctgc tcctgcgtct ggcacacgtg cgtcagatgc	6240
tgccgccacca accgtttacg ctccgtgaga ttcaaacagg cgcttaataa ctgttccata	6300
ttagtcacg cccactggag ctcaggctgg gttttgggg gcaagtaatt ggggatgtag	6360
cactcatcca ccaccttgtt cccgcctccg gcccatttc tggctttgtt gaccgcgaac	6420
cagtttggca aagtccggctc gatccccggg taaattctct gaatcagttt ttccgcaatc	6480
tgactcagga aacgtcccaa aaccatggat ttccacccggg tggtttccac gggcacgtgc	6540
atgtggaagt agctctctcc cttctcaaatt tgcacaaaaga aaagggccctc cggggcccta	6600
ctcacacggc gccattccgt cagaaagtgc cgctgcagct ttcggccac ggtcagggggt	6660
gcctgctcaa ttagatttcg atccatgtca gaatctggcg gcaactccca ttccctctcg	6720
cccaaaaaatgtc tccatctcgta aaaccccggo atggccggctg cgcgttccaa cctcccgctt	6780
acatccatca caatctcgta aaaccccggo atggccggctg cgcgttccaa cctcccgctt	6840
ccaaaatggag accctgcgtg ctcactcggg cttaataacc cagcgtgacc acatgggttc	6900
gcaaaaatgtc gcaaaacact cacgtgaccc ttaatacagg acctccctaa ccctatgacg	6960
taattcacgt cacgactcca cca	6983

<210> SEQ ID NO 38
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 38

caacatacga gccggaagca taaaatgttaa a 31

<210> SEQ ID NO 39
<211> LENGTH: 5334
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 39

ggtagccaaac tccatgtctca acagttccca ggtacagccc accctgcgtc gcaaccagga	60
acagctctac agcttccctgg agcgccactc gcctacttc cgcagccaca gtgcgcagat	120
taggagcgcc acttctttt gtcaatttgc aaacatgtaa aaataatgtt ctagagacac	180
tttcaataaa ggcaaatgtt tttatgttgc cactctcggg tgattattta ccccccaccc	240
tgcgtctgc gccgtttaaa aatcaaggg gttctgcggc gcatcgatat gcccactgg	300
caggacacg ttgcgatact ggtgtttgtt gctccactt aactcaggca caaccatccg	360
cgccagctcg gtgaagttt cactccacag gctgcgcacc atcacaacgc cgtttagcag	420
gtcggggcgc gatatcttgc agtgcgatgtt gggccctccg ccctgcgcgc gcgagttgc	480
atacacaggg ttgcgactt ggaacactat cagcgcggg tggcgcacgc tggccagcac	540

US 12,383,587 B2

141

142

-continued

gctcttgcg gagatcagat ccgcgtccag gtctcccg cgacggagt	600
caacttttgt agctgccttc ccaaaaaggg cgcggtcccc ggctttgagt tgactcgca	660
ccgttagtggc atcaaaaggt gacegtgccc ggtctggcg ttaggataca gcgcctgcat	720
aaaagccttg atctgttaa aagecacctg agectttgcg cttcagaga agaacatgcc	780
gcaagacttg cccgaaaact gattggccgg acaggcccg tcgtcacgc agcacctgc	840
gtcggtgttgcg gagatctgca ccacattcg gccccacccg ttcttcacga tcttggcctt	900
gtagactgc tccttcageg cgcgctgccc gtttgcgtc gtcacatcca tttcaatcac	960
gtgtccctta tttatataa tgcttcgtg tagacactta agctcgccct cgatctcagc	1020
gcagcgggtgc agccacaacg cgacggccgt gggctcgta tgcttgcagg tcacccctgc	1080
aaacgactgc aggtacgcct gcaggaaatcg ccccatcatc gtcacaaagg tcttggct	1140
ggtaagggtc agctgcaacc cgacgggtgtc ctgcgttcagc caggcttgc atacggccgc	1200
cagagcttcc acttggtcag gcagtagttt gaagttcgcc ttagatcgt tatccacgtg	1260
gtacttgtcc atcagcgcgc ggcgcgcctc catgccttc tcccaacgcag acacgatcgg	1320
caacactcagc gggttcatca ccgttaatttc actttccgtc tgcgtggctt cttcccttc	1380
ctcttgcgtc cgcataccac ggcgcactgg gtcgttca ttcagccgc gcactgtgcg	1440
cttacccctt ttgccatgct tgattagcac cgggtgggtt ctgaaaccca ccatttttag	1500
cgccacatct tctctttctt cctcgctgtc cacgattacc tctggtgatg gggggcgtc	1560
gggtttggga gaaggcgct tcttttctt ctggcgca atggccaaat ccggccgcga	1620
ggtcgatggc cgccggctgg gtgtgcgcgg caccagcgcg tcttgcgtatg agtcttcctc	1680
gtcctcggac tcgatacgcc gcctcatccg ctttttggg ggccgcgggg gaggccgcgg	1740
cgacggggac ggggacgaca cgtctccat ggtgggggaa cgtcgcgcgc accgcgtcc	1800
gcgctcgcccc gtggttcgc gctgctcctc ttcccgactg gccatttcct tctcttatag	1860
cgagaaaaag atcatggagt cagtcgagaa gaaggacagc ctaaccgc cctctgagtt	1920
cgccaccacc gcctccaccg atgcgcacaa cgcccttacc accttcccg tcgaggcacc	1980
cccgctttag gaggaggaag tgattatcga gcaggaccca gttttgtaa gcgaagacga	2040
cgaggaccgc tcagtagccaa cagaggataa aaagcaagac caggacaacg cagaggcaaa	2100
cgaggaacaa gtccggcgcc gggacgaaag gcatggcgac tacctatgt tgggagacga	2160
cgtgctgttgc aagcatctgc agcgccagtg cgccattatc tgcgacgcgt tgcaagacgc	2220
cagcgtatgtc cccctcgcca tagcgatgt cggcccttgc tacgaacgcc accttatttc	2280
accgcgcgta ccccccaaac gccaagaaaa cggcacatgc gagcccaacc cgccgcctcaa	2340
tttctacccc gtatggccg tgccagatgt gcttgcacc tacacatct tttccaaaa	2400
ctgcaagata cccctatcct gccgtgccaa ccgcagccga gggacaagc agctggcctt	2460
ggggcagggc gctgtatac ctgatatcgc ctgcgtcaac gaagtgcacaa aaatcttga	2520
gggtcttggaa cgccgacgaga agcgccgcggc aaacgctctg caacaggaaa acagcgaaaa	2580
tgaaagtca tctggagtgt tggtggacta cgagggtgc aacgcgcgc tagccgtact	2640
aaaacgcgcg acgcgacgtca cccactttgc ctacccggca cttAACCTAC cccccaaggt	2700
catgagcaca gtcgtatgt agctgtatgc ggcgcgtgcg cagccctgg agagggatgc	2760
aaatggcaaa gaacaaacag aggaggccct acccgcaatggc acgtacgcgc	2820
ctggcttcaa acgcgcgacgc ctgcgcactt ggaggagcga cgccaaactaa tgatggccgc	2880
agtgcgtt accgtggacgc ttgagtgcat gcagcggttc tttgctgacc cggagatgca	2940

-continued

gcgcaagcta gaggaaacat tgcactacac ctttcgacag ggctacgtac gccaggcctg	3000
caagatctcc aacgtggagc tctgcaacct ggtctcctac ttggaaattt tgcacaaaaa	3060
ccgccttggg caaaaacgtgc ttcatccac gctcaagggc gaggcgccg cgcactacgt	3120
ccgcgactgc gtttacttat ttctatgcta cacctggca acggccatgg gcgttggca	3180
gcagtgcgg gaggagtgca acctcaagga gctgcagaaa ctgctaaagc aaaacttcaa	3240
ggacctatgg acggccatca acgagcgctc cgtggccgccc cacatggccg acatcattt	3300
ccccgaacgc ctgcttaaaa ccctgcaaca gggtctgcca gacttacca gtcaaagcat	3360
gttgcagaac tttagaact ttatccatgca ggcgtcaggaa atcttgcggcc ccacctgctg	3420
tgcacttcct agcgactttg tgccattaa gtaccgcgaa tgccctccgc cgcttgggg	3480
ccactgtcac cttctgcggc tagccaaacta ccttgcctac cactctgaca taatggaa	3540
cgtgagcggt gacgggtctac tggagtgtca ctgtcgctgc aacctatgca ccccgacccg	3600
ctccctgggtt tgcaattcgc agctgcttaa cgaaagtcaa attatcggtt cctttgagct	3660
gcaggggtccc tcgcctgacg aaaagtccgc ggctccgggg ttgaaactca ctccgggct	3720
gtggacgtcg gtttacccgc gcaaatttgtt acctgaggac taccacgccc acgagattag	3780
gttctacgaa gaccaatccc gcccggccaa tgccggagtt accgcctgca tcattaccca	3840
gggcacatt cttggccaat tgcaagccat caacaaagcc cgccaaagagt ttctgctacg	3900
aaagggacgg ggggtttact tggaccccca gtccggccgag gagctcaacc caatcccc	3960
gccggcccgac ccctatcgc agcagccgac ggccttgct tccaggatg gcacccaaaa	4020
agaagctgca gtcggcccg ccacccacgg acgaggagga atactggac agtcaggcag	4080
aggagggtttt ggacgaggag gaggaggaca tggatggaa ctggggagac cttagcagg	4140
aagcttccga ggtcgaagag gtgtcagaacg aaacaccgtc accctcggtc gcattccct	4200
cgccggcgcc ccagaaatcg gcaaccgggtt ccagcatggc tacaacctcc gtcctcagg	4260
cgccggccgc actgcccgtt cgccgacccaa accgtagatg ggacaccact ggaaccagg	4320
ccggtaagtc caagcagccg ccggcggtt cccaaagagca acaacagcgc caaggctacc	4380
gctcatggcg cgggcacaag aacgcctatg ttgttgcgtt gcaagactgtt gggggcaaca	4440
tctccttcgc ccggcgctt cttctctacc atcaccgggtt ggcctccccc cgtaacatcc	4500
tgcattacta ccgtcatctc tacagcccat actgcacccgg cgccagcgcc agcggcagca	4560
acagcagccg ccacacagaa gcaaaaggca ccggatagca agactctgac aaagcccaag	4620
aaatccacag cggcgccagc agcaggagga ggagcgctgc gtctggccgca acacgaaacc	4680
gtatcgaccc gcgagcttag aaacaggatt tttccactc tggatgtat atttcaacag	4740
agcagggggcc aagaacaaga gctgaaaata aaaaacagggt ctctgcgtac cctcaccgc	4800
agctgcgtgt atcacaatggc cgaagatcg cttccggccgca cgctggaaaga cgccggagct	4860
ctcttcagta aatactgcgc gctgactttaa aaggacttagt ttccgcgcctt ttctcaaatt	4920
taagcgcgaa aactacgtca tctccagccg ccacacccgg cgccagcacc tgcgtcagc	4980
gccattatga gcaaggaaat tcccaacgccc tacatgtggaa gttaccagcc acaaattggaa	5040
cttgcggctg gagctgccc agactactca acccgataaa actacatgag cgccggaccc	5100
cacatgatata cccgggtcaa cggatccgc gcccacccgg accgaattctt cttggaaacag	5160
gcccgttata ccaccacacc tcgtaataac cttatcccc gtagttggcc cgctgcctg	5220
gtgttaccagg aaagtccgc tccaccact gtggtaacttc ccagagacgc ccaggccgaa	5280

US 12,383,587 B2

145**146**

-continued

gttcagatga ctaactcagg ggccgcagctt gcggggggct ttcgtcacag ggtg	5334
--	------

```

<210> SEQ_ID NO 40
<211> LENGTH: 3201
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 40

tacactaaac ggtacacagg aaacaggaga cacaactcca agtgcatact ctatgtcatt      60
ttcatggac tggctctggcc acaactacat taatgaaata ttgccacat cctcttacac      120
ttttcatac attgcccagaag aataaagaat cgtttgtt atgtttcaac gtgttattt      180
ttcaattgca gaaaatttca agtcattttt cattcagtag tatagccccca ccaccacata      240
gttatacag atcaccgtag cttaatcaaa ctcacagaac cctagtattc aacctgcccac      300
ctccctccca acacacagag tacacagtcc ttctccccc gttggccta aaaagcatca      360
tatcatgggt aacagacata ttcttaggtt ttatattcca cacggtttcc tgtcgagcca      420
aacgctcata agtgcattttt ataaactccc cgggcagctc acttaagtcc atgtcgctgt      480
ccagctgtg agccacaggg tgctgtccaa cttgcgggtt cttacgggc ggcgaaggag      540
aagtccacgc ctacatgggg gttagtcat aatcgtcat caggataggg cggtggtgct      600
gcagcagcgc gcgaataaac tgctgccccc gcegctccgt cctgcaggaa tacaacatgg      660
cagtggctc ctcagegat attcgcaccc cccgcagcat aaggcgcctt gtcctccggg      720
cacagcagcg caccctgate tcacttaat cagcacagta actgcagcac agcaccacaa      780
tattgttcaa aatcccacag tgcaaggcgc tttatccaa gctcatggcg gggaccacag      840
aacccacgtg gccatcatac cacaagcga ggttagattaa gtggcgaccc ctcataaaca      900
cgctggacat aaacattacc tctttggca ttttgcattt caccacctcc cggtaccata      960
taaacctctg attaaacatg ggcgcattca ccaccatccc aaaccagctg gccaaacct      1020
gcccgcggc tataactgc agggAACGG gactggaaaca atgacagtgg agagccagg      1080
actcgtaacc atggatcatc atgtcgatc tgatatcaat gttggcacaa cacaggcaca      1140
cgtgcataca cttccctcagg attacaagct cttcccgctg tagaaccata tcccaggaa      1200
caacccattc ctgaatcagc gtaaatccca cactgcagg aagacctcgc acgttaactca      1260
cgttgtcat tttttttttt ttatccaa agattatcca aacccatccaa ggcggccctg      1320
cgccgggttc tttttttttt ttatccaa agattatcca aacccatccaa ggcggccctg      1380
accgagatcg tttttttttt ttatccaa agattatcca aacccatccaa ggcggccctg      1440
ctgaagccaa accaggatcg ggcgtgacaa acatgcgtc gtcgtccggc tccggccctg      1500
gatcgctctg tttttttttt ttatccaa agattatcca aacccatccaa ggcggccctg      1560
gcttcgggtt ctatgtaaac tccttcatgc gccgctgccc tgataacatc caccacccca      1620
gaataagccaa caccacccaa acctacacat tcgttctgcg agtcacacac gggaggagcg      1680
ggaaagatcg gaagaaccat gttttttttt ttatccaa agattatcca aacccatccaa      1740
atgaagatct attaagtgaa cgcgtccccc tccgggtggcg tggtcaaact ctacagccaa      1800
agaacacata atggcatttg taatgttg cacaatggct tccaaaggc aacccgtccct      1860
cacgtccaaag tggacgtaaa ggctaaaccc ttcagggtga atctcctcta taaacattcc      1920
agcacccatca accatgccc aataattctc atctcgccac cttctcaata tatctctaag      1980

```

US 12,383,587 B2

147**148**

-continued

caaataccga atattaagtc cggccattgt aaaaatctgc tccagagcgc cctccacctt	2040
cagcctcaag cagcgaatca tgattgcaaa aattcagggtt cctcacagac ctgtataaga	2100
ttcaaaaagcg gaacattaac aaaaataccg cgatcccgta ggtcccttcg cagggccagc	2160
tgaacataat cgtcagggtc tgcacggacc agegcggcca cttccccgcc aggaaccatg	2220
acaaaaagaac ccacactgtat tatgacacgc atactcgtag ctatgtaac cagcgtagcc	2280
cccgatgtaaat cttgttgcat gggeggcgat ataaaatgca aggtgtgtct caaaaaatca	2340
ggcaaaggct cgcgcaaaaa agaaagcaca tcgttagtcat gtcgtgcag ataaaggcag	2400
gttaagctccg gaaccaccac agaaaaagac accattttc tctcaaacat gtctgcgggt	2460
ttctgcataa acacaaaata aaataacaaa aaaacattt aacattagaa gcctgtctta	2520
caacaggaaa aacaaccctt ataaggcataa gacggactac gccatgcgcg gcgtgacccgt	2580
aaaaaaaaactg gtcaccgtga ttaaaaagca ccaccgacag ctccctcgat atgtccggag	2640
tcataatgtaa agactcggtaa aacacatcgtt gttgattcac atcggtcagt gctaaaaagc	2700
gaccgaaata gcccggggga atacataccgc gcaggcgtag agacaacatt acagccccca	2760
taggaggtt aacaaaatta ataggagaga aaaacacata aacacctgaa aaaccctcct	2820
gcctaggcaa aatagcaccc tcccgttcca gaacaacata cagcgttcc acagcggcag	2880
ccataacagt cagccttacc agtaaaaaag aaaacctatt aaaaaaacac cactcgacac	2940
ggcaccagct caatcgtca cagtgtaaaa aaggggcaag tgcagagcga gtatatacg	3000
gactaaaaaa tgacgttaacg gttaaagtcc aaaaaaaaaa cccagaaaaac cgacgcgaa	3060
cctacgcccga gaaacgaaag cccaaaaacc cacaacttcc tcaaattcgtc acttccgttt	3120
tcccacgtta cgtaacttcc cattttaa aactacaat tcccaacaca tacaagttac	3180
tccgccttaa aacctacgtc a	3201

<210> SEQ ID NO 41
 <211> LENGTH: 6896
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 41

ctagagtcga cctgcaggca tgcaagcttgcgtatcat ggtcatacgat gtttctgttgcgtatcat	60
tgaaattgtt atccgtcac aattcccacac aacatacgttgcgtatcat ggtcatacgat gtttctgttgcgtatcat	120
gcctgggggttgcgtatcat gagctaactc acattaatttgcgtatcat ggtcatacgat gtttctgttgcgtatcat	180
ttccagtcgg gaaacctgttcgtatcat ggtcatacgat gtttctgttgcgtatcat	240
ggcgggttgcgtatcat ggtcatacgat gtttctgttgcgtatcat	300
gttcgggtgcgtatcat ggtcatacgat gtttctgttgcgtatcat	360
tcaggggata acgcaggaaa gaacatgttgcgtatcat ggtcatacgat gtttctgttgcgtatcat	420
aaaaaggccgcgtatcat ggttccat aggctccgcgcgtatcat ggtcatacgat gtttctgttgcgtatcat	480
aatcgacgtatcat ggtcatacgat gtttctgttgcgtatcat	540
ccccctggaa gtccttcgtatcat ggttccat aggctccgcgcgtatcat ggtcatacgat gtttctgttgcgtatcat	600
tccgccttcgtatcat ggttccat aggctccgcgcgtatcat ggtcatacgat gtttctgttgcgtatcat	660
agttcggtgttgcgtatcat ggttccat aggctccgcgcgtatcat ggtcatacgat gtttctgttgcgtatcat	720
gaccgctgcgtatcat ggttccat aggctccgcgcgtatcat ggtcatacgat gtttctgttgcgtatcat	780

US 12,383,587 B2

149**150**

-continued

tcgcccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt aggccgtgct	840
acagagttct tgaagtggtg gcctaactac ggctacacta gaaggacagt atttggtatac	900
tgcgctctgc tgaagccagt tacttcgga aaaagagttg gtagctctg atccggcaaa	960
caaaccaccc ctggtagcggt tggtttttt gtttgcaga acgagattac ggcgcagaaaa	1020
aaaggatctc aagaagatcc tttgatctt tctacggggt ctgacgctca gtggacgaa	1080
aactcacgtt aagggatttt ggtcatgtga ttatcaaaaa ggatcttacat ctagatcctt	1140
ttaaattaaa aatgaagttt taaatcaatc taaagtataat atgagtaaac ttggctgac	1200
agtttagaaaa actcatcgag catcaaatga aactgcaattt tattcatatc aggattatca	1260
ataccatatt tttgaaaaag ccgtttctgt aatgaaggag aaaactcacc gaggcagttc	1320
cataggatgg caagatcctg gtatcggtct gcgatcccgta ctcgtccaac atcaatacaa	1380
cctattaatt tccccctcgta aaaaataagg ttatcaagtg agaaatcacc atgagtgacg	1440
actgaatccg gtgagaatgg caaaagttt tgcatttctt tccagacttg ttcaacaggc	1500
cagccattac gtcgtcatc aaaatcactc gcatcaacca aaccgttatt cattcgtgat	1560
tgcgcctgag cgagacgaaa tacgcgtatcg ctgtttaaaag gacaattaca aacaggaatc	1620
gaatgcaacc ggccgcaggaa cactgccago gcatcaacaa tattttcacc tgaatcagga	1680
tattcttcta atacctggaa tgctgttttccagggatcg cagttgttag taaccatgca	1740
tcatcaggag tacggataaa atgcttgcgtatcg gtcggaaagag gcataaaatc cgtcagccag	1800
ttagtgtctga ccatctcatc tgtaacatca ttggcaacgc taccttgcc atgtttcaga	1860
aacaactctg ggcgcacccggg cttccctatac aatcgataga ttgtcgacc tgattgccc	1920
acattatcgc gagcccatat atacccatataatcagcat ccatgttgaa atttaatcgc	1980
ggccttagagc aagacgtttc ccgttgaata tggctcatac tttccctttt tcaatattat	2040
tgaagcattt atcagggtta ttgtctcatg agcggatatac tatttgaatg tatttagaaa	2100
aataaaacaaa taggggttcc ggcacacattt ccccgaaaag tgccacctga cgtctaagaa	2160
accattatataatcagacattt aacctataaa aataggcgta tcacgaggcc ctttcgtctc	2220
gcccgtttcg gtgatgcgg tgaaaacctc tgacacatgc agctcccgga gacggtcaca	2280
gtttgtctgt aagcggatgc cgggagcaga caagcccgta agggcgcgtc agcgggtgtt	2340
ggccgggtgtc ggggctggct taactatcg gcatcagacgc agattgtact gagagtgcac	2400
catatgcgggt gtgaaatacc gcacagatgc gtaaggagaa aataccgcatt caggccat	2460
tcgcccattca ggctgcgca ctgttggaa gggcgtatcg tgccggccctc ttgcgtat	2520
cgccagctgg cgaaaggggg atgtgctgca aggcgattaa gttggtaac gccagggttt	2580
tcccaagtcac gacgttgtaa aacgcacggcc agtgaatttc agctcggtac ccgttagccat	2640
ggaaactaga taagaaagaa atacgcagag accaaagttc aactgaaacg aattaaacgg	2700
tttatttattt aacaagcaat tacagattac gagtcaggtt tctgggtccca atggggccgg	2760
gttcactata tacacccatca gtatcaacag caaattcaac attattagac ttgtatagt	2820
tggaaagtgtt ctggatctcc ggggtccago gcttgcgtgtt tccctctgc agctccact	2880
cgatctccac gctgacttgg ccagtagaat actgggtgtat gaaagagtcc agcttgcct	2940
tgttgaaggc cggtggagga tccgcaggta caggtgtgtt tttgatgagg atctgaggag	3000
ggccgggtgtt cattccaaac cctccatca gcccggaaagg gtgaaagtgc ccgtccgtgt	3060
gaggaattttt ggcccaatg ggtccttgcg ggtacacatc tctgtctgc caaaccatac	3120
ccggaaagtat tccttggttt tgaacccago cggctctgcgc ctgtgtttgg gcactctgg	3180

ggtttgtggc cacttgcata taggactccg ttgttacccgg gtttagtagtt ttaatttctt 3240
 ctgcgttggt tatcatgact ttgtccgcata ccacgttgc tcttccaggt ccttgcgttgc 3300
 caaaaatata agatccagac aaaggaaaga aacggtcctc tcttcttgc tggctggcca 3360
 tagcagggtcc aggattcata aagctattac gtccatttagag agcccaagaa gaagctccag 3420
 gccaagcaaa ttgcgttgc ttgttttagag tcacagtggt tgagacacgt tggtgtcggt 3480
 agctgggtcc aggtatgttag ttcttccct ggacagccat gttgctgggt ccggccacac 3540
 tgaatttttag cgttgttgc ttctgtccag aaccgttaat agtcttgcag agatagtaca 3600
 agtattggtc gatgagtgta ttcatgttc ggtccaggct ttggctgtga gcgttagctgc 3660
 tatggaaagg tacgttctca aactcgtagc tgaactggaa gttgttaccc gttcttagca 3720
 ttgcgcacgg gaaatattcc aggccgtaaa aggacgaaac accccacggcc tggcttccat 3780
 cattaagcgt cagatacccg tactgaggaa tcatgaaaac gtccgttggg aacggccggga 3840
 ggcagccctc gtgagccgac ccggcacgt acgggagctg atagtcgttgc tccgtgaaga 3900
 cctggaccgt gctggtaagg ttattggcga tggtcttgc tccattgttgc tccgttaacct 3960
 ctttgacctg aatgttgcag agcttgcgtt tgagtcgttgc tggccggaaat ccccgatgt 4020
 tggatgttgc tggctgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4080
 aataccccca ggggggtgtg tagccgttgc tggccgttgc atttgcgttgc cctccagatg 4140
 tgcgttgcgat gatttgcgttgc tagaggtgtat tggatgttgc gggcaggggcc cagggttccgg 4200
 tgcgttgcgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4260
 aactaccaccc tccatcgccat cttcggttat tgcgttgc tggatgttgc cccctgttgc 4320
 ccattgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4380
 ctgggactga ctctgtgtcg ccagtctgc tggatgttgc tggatgttgc cccctgttgc 4440
 cacccgattt gcaataaccc gcccggggat cccgttgc tggatgttgc tggatgttgc 4500
 tcttcttcc agggccgttgc ttagccgttgc cctcaaccat accaagggat tcaagaagcc 4560
 tcttttgcgat gctggccgttgc tggatgttgc tggatgttgc tggatgttgc 4620
 gccgctctg tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4680
 tggatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4740
 gctcccccgtt gtcgttgc tggatgttgc tggatgttgc tggatgttgc 4800
 ctggatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4860
 cccaccactc gcaatttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4920
 cggcagccat acctgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 4980
 ttgaccat gcaacttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5040
 cacatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5100
 cactctaaac agtcttttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5160
 tgcattgtc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5220
 tggatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5280
 gactcgccat cccgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5340
 gtcaccat tggatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5400
 cggaaaaatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5460
 gtgagttcaat tggatgttgc tggatgttgc tcaactgggtt agaagtggca gtggaaatctg ttgaagtcaa 5520

-continued

ttcccgtaa tcacggcgca catgttggtg ttggagggtga cgatcacggg agtcgggtct	5580
atctgggccc aggacttgcata ttcttggtcc acgcgcaccc tgcgttccccc gagaatggct	5640
ttggccgact ccacgaccc ttcccccctt cccaccagat caccatcttgc	5700
tgcacacagt cgttgaaggg aaagtctca ttggtccagg ttacgcaccc gtagaagggc	5760
acagtgtggg ctatggcctc cgcatgttg gtcttcccg tagttgcagg cccaaacagc	5820
cagatggtgt tccttggcc gaacttttc gtggcccatc ccagaaagac ggaagccgca	5880
tattggggat cgtaccggtt tagttccaaa atttataaa tccgattgtt ggaaatgtcc	5940
ttcacgggct gctggccac caggttagtcg gggggcggtt tagtcaggct cataatctt	6000
cccgccattgtt ccaaggcagc cttgatttg gaccgcagat tggaggccgc attgaaggag	6060
atgttatgagg cctggccctc ctggatccac tgcttctccg aggtaatccc cttgtccacg	6120
agccacccga ccagctccat gtacctggct gaagtttttgc atctgatcac cggcgcatca	6180
gaattgggat tctgattctc tttgttctgc tcctgcgttgc ggcacacgtt cgtcagatgc	6240
tgccgccacca accggttacg ctccgtgaga ttcaaacagg cgcttaataa ctgttccata	6300
ttagtccacg cccactggag ctcaggctgg gttttgggg gcaagtaatt ggggatgttag	6360
cactcatcca ccaccttgc cccgcctccg ggcatttc tggctttgttgc gaccgcgaac	6420
cagtttggca aagtccggc gatccccggg taaattcttgc gaatcgtttt ttcgcgaatc	6480
tgactcagga aacgtcccaa aaccatggat ttccacccggg tggtttccac gggcacgtgc	6540
atgttggaaatg agctctctcc ctctcaat tgcacaaaga aaagggccctc cggggccctt	6600
ctcacacggc gccattccgt cagaaagtgc cgctgcagat tctcgccac ggtcagggtt	6660
gcctgctcaa tcagattcag atccatgtca gaatctggcg gcaactccca ttccctctcg	6720
gccacccaggat tcacaaagct gtcagaaatgc ccgggcagat gtcgtcaag gtcgtgggg	6780
acctaataatca caatctcgta aaacccggc atggcggctg cgccagatc aagttcttat	6840
actttctaga gaataggaac ttccggatag gaacttctga ttccgggg gatcca	6896

<210> SEQ ID NO 42
 <211> LENGTH: 8469
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (4155)..(4254)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (4155)..(4254)
 <223> OTHER INFORMATION: This region may encompass 60-100 nucleotides

<400> SEQUENCE: 42

tagggaaata ggccagggtt tcaccgtAAC acgccacatC ttgcgtatAT atgtgtAGAA	60
actgcccggaa atcgtcgtgt gcactcatgg aaaacgggtt aacaagggtt aacactatCC	120
catatcacca gtcaccgtc ttccattGCC atacggAACT ccggatgagC attcatcagg	180
ccggcaagaa tgtgaataaa ggccggataa aacttgcgtt tatttttctt tacggtctt	240
aaaaaggccg taatatccag ctgaacggtc tggttatagg tacattgagC aactgactGA	300
aatgcctcaa aatgttctt acgatgccat tggatataat caacgggtt atatccagtg	360
attttttctt ccattttttt ttccctcttt agaaaaactC atcgagcatC aaatgaaact	420

US 12,383,587 B2

155**156**

-continued

gcaatttatt catatcagga ttatcaatac catattttg aaaaagccgt ttctgtaatg	480
aaggagaaaa ctcacccgagg cagttccata ggatggcaag atcctggtat cggctcgca	540
ttcccgactcg tccaaacatca atacaaccta ttaatttccc ctcgtcaaaa ataaggttat	600
caagtgagaa atcaccatga gtgacgactg aatccggta gaatggcaaa agtttatgca	660
tttcttcca gacttgtca acaggccago cattacgctc gtcatcaaaa tcactcgcat	720
caaccaaacc gttattcatt cgtgattgctc cctgagcgag gcgaaatacg cgatcgctgt	780
taaaaggaca attacaaca ggaatcgagt gcaaccggcg caggaacact gccagcgcat	840
caacaatatt ttcacctgaa tcaggatatt cttaatac ctggAACGCT gttttccgg	900
ggatcgcagt ggtgagtaac catgcatcat caggagtacg gataaaatgc ttgatggctg	960
gaagtggcat aaattccgtc agccagtttta gtctgaccat ctcatctgta acatcatgg	1020
caacgcgtacc tttgccatgt ttcagaaaca actctggcgatc atcgggcttc ccatacaagc	1080
gatagattgt cgcacctgat tgcccgacat tatcgcgagc ccatttatac ccataataat	1140
cagcatecat gttgaaattt aatcgccggcc tcgacgtttc ccgttgaata tggctcattt	1200
tttttccct ctttaccaat gcttaatcg tgaggcacct atctcagcgatc tctgtctatt	1260
tcgttcatcc atagttgcct gactccccgt cgtgttagata actacgatac gggaggcgtt	1320
accatctggc cccagcgctg cgatgatacc gcgagaacca cgctcacccgg ctccggattt	1380
atcagcaata aaccagccag ccggaaaggcc cgagcgccaga agtgggtccctg caactttatc	1440
cgcctccatc cagtctatta atttttgcgg ggaagctaga gtaagtagtt cgccagttaa	1500
tagtttgcgc aacgttggc ccatcgctac aggcatcgatc gtgtcacgtc cgctgtttgg	1560
tatggcttca ttcagctccg gttcccaacg atcaaggcgatc gttacatgtat ccccatgtt	1620
gtgcacgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat ggcagcaactg	1680
cataattctc ttactgtcat gccatccgtat agatgctttt ctgtgactgg tgagtactca	1740
accaagtcat tctgagaata gtgtatgcgg cgaccgaggat gctcttgcgg ggcgtcaata	1800
cgggataata ccgcgcacaca tagcagaact ttaaaagtgc tcatcattgg aaaacgttct	1860
tcggggcgaa aactctcaag gatcttaccg ctgttgcgat ccagttcgat gtaaccact	1920
cgtgcaccca actgatcttc agcatctttt actttcacca gctttctgg gtgagcaaaa	1980
acagggaggc aaaatgcgcgca aaaaaaggga ataaggcgatc cacggaaatg ttgaataactc	2040
atattcttcc ttttcaataa ttattgaagc atttatcagg gttattgtct catgagcgat	2100
tacatatttg aatgttattta gaaaaataaa caaatagggg tcagtgatc aaccaattaa	2160
ccaattctga acattatcgatc gagccattt atacctgaat atggctatac acaccccttgc	2220
tttgcctggc ggcagtagcg cgggtggccc acctgacccc atgcccgaact cagaagtgaa	2280
acgcccgtac gcccgtggta gtgtggggac tccccatgcg agagtagggc actgccaggc	2340
atcaaataaa acgaaaggct cagtcgaaag actgggcctt tccggccggc taattgaggg	2400
gtgtcgccct tattcgactc ggggtcgag cagcagctgc gctcgctc gctactgag	2460
ggccggccggg caaagcccg gctgtggggcc acctttggatc gcccggccctc agtgcgcgag	2520
cgagcgccca gagagggaggat ggggtttaa ttaaacgcgt ttacataact tacggtaat	2580
ggcccgccctg gctgaccggcc caacgacccc cgccattga cgtcaataat gacgtatgtt	2640
cccatagtaa cgccaaatagg gactttccat tgacgtcaat ggggtggacta tttacggtaa	2700
actgcccact tggcagttaca tcaagtgtat catatgccaat gtaacccccctt tattgacgtc	2760
aatgacggta aatggcccgcc ctggcattat gcccaggatca tgaccttgcgatc ggactttcc	2820

acttggcagt acatctacgt attagtcatc gctattacca tggtgatgcg gttttggcag 2880
 tacatcaatg ggccgtggata cgccgtttgac tcacggggat ttccaagtct ccaccatt 2940
 gacgtcaatg ggagtttgtt ttggcaccaa aatcaacggg actttccaaa atgtcgtaac 3000
 aactccgccc cattgacgca aatggggcgtt aggcgtgtac ggtgggaggt ctatataaggc 3060
 gcgccgaact gaaaaaccag aaagttaact ggtaagttt gtcttttgtt cttttatttc 3120
 aggtccccggaa tccgggtgtt gtgc当地 aagaactgtt ctcagtgaa tggtgcctt 3180
 acttcttaggc ctgtacggaa gtgttacttc tgctctaaa gtcctgcag ggaattcgcc 3240
 accatggtga gcaaggcgcga ggagctgtt accgggggtgg tgcccatctt ggtcgagctg 3300
 gacggcgacg taaacggcca caagttcage gtgtccggcg agggcgaggg cgatgccacc 3360
 tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gcccctggccc 3420
 accctcgtga ccaccctgac ctacggcggt cagtgc当地 gcccgttacc cggaccatg 3480
 aacgcacgacg acttcttcaa gtccggcatg cccgaaggct acgtccagga ggc当地acc 3540
 ttcttc当地agg acgacggcaa ctacaagacc cggcccgagg tgaagttcga gggcgacacc 3600
 ctgggtgaacc gcatcgagct gaaggcgc当地 gacttcaagg aggacggcaa catctgggg 3660
 cacaagctgg agtacaacta caacagccac aacgtctata tcatggocga caagcagaag 3720
 aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtc当地ctc 3780
 gccgaccact accagcagaa caccggccatc ggccgacggcc cctgtctgtt gccc当地aac 3840
 cactacctga gcacccagtc cgccctgago aaagacccca acgagaaggc cgatcacatg 3900
 gtccctgtgg agtccgtgac cgccgggggg atcactctcg gcatggacga gctgtacaag 3960
 taatagacta gtgtcgacag atctttaaa aaacctccca cacaattgtt gttt当地act 4020
 tgtttattgc agtttataat gttt当地aaat aaagcaatag catcacaat tt当地acaata 4080
 aagcatttt tt当地actgcat tcttagttgtt gttt当地ccaa actcatcaat gtatcttac 4140
 atgtctgtttt aaacnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4200
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnattaa 4260
 ataggaaccc ctagtgtatgg agttggccac tccctctcg cgc当地ctgct cgtc当地ta 4320
 ggccggggcga ccaaaggctg cccgacgccc gggctttgc当地 cggccggcct cagtgagcga 4380
 gcgagcgc当地 agagactata categatgtg agttcgccgg tggctgggg gccc当地ggct 4440
 ggccggccccc cccgaccccg cttacccgacc ttggccgttcc cgggtcttgc当地 cgtc当地tatg 4500
 ggggcagggtt agctgttccc cgcaaggaga gctcaaggctc agcgctcgga cctggcggag 4560
 cccccc当地ccc aggctgtggc gccc当地tgca gctccggccct tggccggccca tctgcccgg 4620
 gccc当地ctcc cctagtcccc agaaaaggaa ggtccctactt cccgccc当地gag atccc当地accc 4680
 ggacccctag gtggggacg cttt当地ttcc tttccggctc tggccgggtca cgttgc当地cag 4740
 aggagccccc cccccc当地ggc cttccggccacc gcaggccccc ggatgttagt ggc当地ggcggg 4800
 tgc当地ccctg tccggatgtc ggc当地ctgccc tagagccggcc gccatgttgc aaccgggaaag 4860
 gaaaatgtatgg ggc当地ccgtt agggaaaggctt gccggctactt aaccctgc当地 tccctgc当地ctcg 4920
 atgggtggag tggc当地gtgg cggggaaaggctc aggtggagcg aggctagctg gccc当地atttc 4980
 tccctccgggtt gatgttttc ctagattattt ctctggtaaa tcaaaggaaatgg gggttt当地atgg 5040
 aggtcccttcc tggccccc当地tcc cccgacagg gttgggtggct gtc当地atggt gccaaggccgg 5100
 gagaaggctga gtc当地atggta gttggaaaag gacatttccca cccgcaaaaatg gccc当地ctgg 5160

US 12,383,587 B2

159**160**

-continued

tggtgtggcccc	ttcctgcagc	gccggctcac	ctcacggccc	cggcccttccc	ctgccaggct	5220
agcggttacc	cgaccccaaa	ggcaggctg	taaatgtcac	cgggaggatt	gggtgtctgg	5280
gccccteggg	gaacctgccc	ttctccccat	tccgtttcc	ggaaaccaga	tctcccaccc	5340
caccctggtc	tgaggttaaa	tatagtctgt	gacctttctg	tagctgggg	cctgggctgg	5400
ggctctctcc	catcccttct	ccccacacac	atgcactac	ctgtgtctcc	actcctgtatt	5460
tctggaaaag	agcttaggaag	gacaggcaac	ttggcaaatac	aaagccctgg	gactaggggg	5520
ttaaaataca	gttccccctc	ttcccacccg	ccccagtctc	tgtccctttt	gtaggaggga	5580
cttagagaag	gggtgggctt	gccctgtcca	gttaatttct	gacctttact	cctgcccctt	5640
gagttttagt	atgctgagtg	tacaagcggt	ttctccctaa	agggtgcagc	tgagctaggc	5700
agcagcaagc	attctctgggg	tggcatagtg	gggtgggtgaa	taccatgtac	aaagtttg	5760
cccagactgt	gggtggcagt	gccccacatg	gccgcttctc	ctggaaaggc	ttcgtatgac	5820
tgggggtgtt	gggcagccct	ggagccttca	gttgcaagca	tgccttaagc	caggccagcc	5880
ttggcaggaa	gctcaaggaa	gataaaatcc	aaccttctgg	gccctctgg	gggttaaggag	5940
atgctgcatt	cgccctctta	atggggaggt	ggccttagggc	tgctcacata	ttctggagga	6000
gcctccccctc	ctcatgcctt	cttgccttct	gtctctttagg	catgcaaaag	agtcaataaa	6060
gggcgcacaca	aaatttattc	taaatgcata	ataaatactg	ataacatctt	atagttgt	6120
ttatattttg	tattatcggt	gacatgtata	attttgatata	aaaaaactga	ttttcccttt	6180
attattttcg	agatttatct	tcttaattct	ctttaacaaa	ctagaaatata	tgtatataca	6240
aaaaatcata	aataatagat	gaatagttt	attataggtt	ttcatcaatc	aaaaaagcaa	6300
cgtatcttat	ttaaagtgcg	ttgctttttt	ctcatttata	aggttaataa	attctcatat	6360
atcaagcaaa	gtgacaggcg	cccttaataa	ttctgacaaa	tgctcttccc	ctaaaactccc	6420
cccataaaaaa	aacccgcccga	agcggggtttt	tacgttattt	gcggattaac	gattactcgt	6480
tatcagaacc	gcccaggggg	cccggagctt	agactggccg	tcgttttaca	acacagaaag	6540
agttttaga	aacgcaaaaaa	ggccatccgt	caggggcctt	ctgcttagtt	tgatgcctgg	6600
cgatccccta	ctctcgccctt	ccgcttccctc	gctcaactgac	tcgtctgcgt	cggtcggtcg	6660
gtgcggcga	gcggtatcag	ctcaactaaa	ggcggtataa	cggttatcca	cagaatcagg	6720
ggataaacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaaa	6780
ggccgcgttg	ctggcggttt	tccataggct	ccggccccct	gacgagcatc	acaaaaatcg	6840
acgctcaagt	cagagggtggc	gaaacccgac	aggactataaa	agataccagg	cgtttcccc	6900
tggaaagctcc	ctcgctgcgt	ctccgttcc	gaccctgccc	cttaccggat	acctgtccgc	6960
ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	cgctgttaggt	atctcagttc	7020
ggtgttaggtc	gttcgttcca	agctgggctg	tgtgcacgaa	ccccccgttc	agcccgaccg	7080
ctgcgcctta	tccggtaact	atcgcttgc	gttcaacccg	gtaagacacg	acttatcgcc	7140
actggcagca	gccactggta	acaggattag	cagagcgagg	tatgttaggc	gtgtacaga	7200
gttcttgaag	tggtgggctt	actacggcta	cactagaaga	acagtatttg	gtatctcg	7260
tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	tcttgcgtcc	gcaaacaac	7320
caccgctgg	agcggtggtt	tttttggttt	caagcagcag	attacgcgc	gaaaaaaaaagg	7380
atctcaagaa	gatcccttgc	tcttttctac	ggggctgcac	gtcgttgcga	acgacgcgc	7440
cgtaactcac	gttaagggtat	tttggctatg	agcttgcgc	gtcccgtaa	gtcagcgtaa	7500
tgctctgcctt	aggtggcggt	acttgggtcg	atataaagt	gcatcacttc	ttcccgat	7560

US 12,383,587 B2

161

-continued

162

```

cccaactttg tatagagagc cactgcggga tcgtcacgt aatctgcttg cacgttagatc    7620
acataaggac caagcgcggtt ggcctcatgc ttgaggagat tgatgagcgc ggtggcaatg    7680
ccctgcctcc ggtgctcgcc ggagactgcg agatcataga tatacatctc actacgcggc    7740
tgctcaaact tgggcagaac gtaagccgcg agagcgccaa caaccgccttc ttggtcgaag    7800
gcagcaagcg cgatgaatgt cttaactacgg agcaagtcc cgaggtaatc ggagtccggc    7860
tgatgttggg agtaggtggc tacgtcacgg aactcacgcg cgaaaagatc aagagcagcc    7920
cgcatggatt tgacttggc agggccgagg ctacatgtgc gaatgtgcc catacttgag    7980
ccacctaact ttgttttagg gcgactgccc tgctgcgtaa catcggttgc gctccataac    8040
atcaaacatc gaccacgcgc gtaacgcgt tgctgcttgg atgcccggagg catagactgt    8100
acaaaaaaaac agtcataaca agccatgaaa accgcccactg cgccgttacc accgctgcgt    8160
tcggtcaagg ttctggacca gttgcgttag cgcatttttt tttccctccgc ggcgtttacg    8220
ccccggccctg ccactcatcg cagtagtgg ttaatttattt aagcattctg ccgacatggg    8280
agccatcaca gacggcatga tgaacctgaa tcgcccggg catcagoacc ttgtgcctt    8340
gcgtataata ttggccata gtgaaaacgg gggcgaagaa gttgtccata ttggccacgt    8400
ttaaatcaaa actggtgaaa ctcacccagg gattggcgct gacgaaaaac atattctcaa    8460
taaacccctt                                         8469

```

```

<210> SEQ ID NO 43
<211> LENGTH: 116
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

```
<400> SEQUENCE: 43
```

```

cagcagctgc gcgctcgctc gctcaactgag gcccggggg caaagccgg gcgtcgccg     60
acctttggtc gcccggctc agtgagcgag cgagcgcgcg gagagggagt ggggtt           116

```

```

<210> SEQ ID NO 44
<211> LENGTH: 534
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

```

```
<400> SEQUENCE: 44
```

```

ttaggtggcg gtacttgggt cgatataaaa gtgcataact tttcccgta tgcccaactt    60
tgtatagaga ggcactcgcc gatcgctacc gtaatctgct tgcaactaga tcacataagc    120
accaagcgcg ttggcctcat gcttgaggag attgtgagc gcggtggcaa tgccctgcct    180
ccgggtgtcg ccggagactg cgagatcata gatatacgatc tcaactacgcg gctgtcaaa    240
cttggcaga acgtaagccg cgagagcgcc aacaaccgct tttggtcga aggccagcaag    300
cgcgatgaat gtcttactac ggagoaagtt cccgaggtaa tggaggatccg gctgtgttg    360
ggagtaggtg gtcacgtcac cgaactcagc accgaaaaaga tcaagagcag cccgcattgg    420
tttgacttgg tcagggccga gcctcatgt gcaatgtat cccatacttg agccaccaa    480
ctttgtttta gggcgactgc cctgctgcgt aacatcggtt ctgctccata acat            534

```

```
<210> SEQ ID NO 45
```

-continued

```

<211> LENGTH: 11635
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 45

ggtaaccaac tccatgctta acagtcccc ggtacagccc accctgcgtc gcaaccagga      60
acagctctac acgttctgg agcgccactc gcctacttc cgcageccaca gtgcgcagat      120
taggagcgcc actttttttt gtcacttgaa aaacatgtaa aaataatgtta ctaggagaca      180
ctttcaataa aggcaaatgt ttttatgtt acactctcggt gtgattatcc acccccccacc      240
cttgcgtct ggcgcgttta aaaatcaaag gggttctgcc ggcgcacgt atgcgcact      300
ggcagggaca cggtgcgata ctgggtttta gtgcgtccact taaactcagg cacaaccatc      360
cgccgcagct cggtgaagtt ttcactccac aggtgcgcac ccacccacca cgcgtttagc      420
aggtcgccgcg ccgatatctt gaagtgcgcg ttggggcctc cgcgcgtgc ggcgcagtt      480
cgatacacag gggtgcagca ctgaaacact atcagcgcgcg gggtggtgcac gctggccagc      540
acgtcttgtt cggagatcag atccgcgtcc aggttctccg cgttgcgtcag ggcgaacgg      600
gtcaactttt gtagctgcct tccaaaaaag gggtgcgtcc caggcttgcgtt gttgcactcg      660
caccgttagt gcatcagaag gtgaccgtgc cccgtctggg cgttgcgttgc cagcgcctgc      720
atgaaaaggct tgatctgctt aaaagccacc tgagcctttt cgccttcaga gaagaacatg      780
ccgcaagact tgccggaaaa ctgattggcc ggacaggccc cgtcatgcac gcgcacccctt      840
gcgtcggtgt tggagatctg caccacattt cggccccacc gggttctcac gatcttggcc      900
ttgtcttagact gtccttcag cgcgcgtgc cccgtttcg cgttgcgttgc catttcatac      960
acgtgcgttct tatttatcat aatgtcccg ttagacact taagctgcgc ttgcgttca      1020
gcgcageggcgt gcagccacaa cgcgcagccc gtgggtcgat ggtgcgttgc gtttacccct      1080
gcaaacact gcaggtaacgc ctgcaggaaat cgcgcacccatca tcgtcacaatgg tttttttt      1140
ctgggtgaagg ttagctgcac cccgcgtgc tcctcggtt ggcagggttgc gcatacgcc      1200
gccagagctt ccacttggc aggcagtagt ttgaagtttgc ctttagatc gttatccacg      1260
ttgtacttgtt ccatcaacgc gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc      1320
ggcaggtca gcgggtttat cccgtgttgc tcacttccgc ctttgcgttgc ctttccctt      1380
tcctcttgcg tccgcatacc cccgcgcact ggggtgttgc cattcagccgc cccgcacccgt      1440
cgcttacccgc ctttgcgtgc ctttgcgttgc accgggtgggt tggtaaccacc caccatttt      1500
agcgcacat cttcttttc tcctcggttgc tccacgttgc cctctggggta tggcgccgc      1560
tcgggttttttgg gagaggggcg ctttttttttgc tttttggacg caatggccaa atccgcgtcc      1620
gagggtcgatg ggcgcgggttgc ggggtgtgcgc ggcaccagcc catcttgcgttgc ggggttgc      1680
tcgtcctcgcc actcgagacg cccgcctcago cgcgttttttgc ggggcgcgcgcgcgcgcgc      1740
ggcgcacggcg acggggacga cacgttccgc atgggttgggt gacgtgcgcgcgcgcgcgcgc      1800
ccgcgcgttgc ggggtgttttgc ggcgtgttgc tcctccgcgc tggccatttc ctttccat      1860
aggcagaaaa agatcatggt gtcagtcgag aaggaggaca gcttaaccgc cccctttgag      1920
ttcgccacca cccgcctccac cgtatgcgcgc aacgcgcctca ccacccccc cgtcgaggca      1980
cccccgcttgc aggaggagga agtgattatc gacgaggacc caggtttgtt aagcgaagac      2040
gacgaggatc gtcagtcgttcc aacagaggat aaaaagcaag accaggacga cgcagaggca      2100

```

-continued

aacgaggaac aagtccggcg gggggaccac	2160
gacgtgtgt tgaagcatct gcagcgccag tgegccatta tctgcgacgc gttcaagag	2220
cgcagcgatg tgccccctgc catagcggat gtcagccttg cctacgaacg ccacctgttc	2280
tcaccgegcg taccccccac acgccaagaa aacggcacat gcgageccaa cccgegcctc	2340
aacttctacc ccgtatttgc cggtccagag gtgtttgtcca cctatcacat cttttccaa	2400
aactgcaaga tacccctatc ctgcgcgtgc aaccgcagcc gagcggacaa gcagctggcc	2460
ttgcggcagg gcgctgtcat acctgatata gcctcgctcg acgaagtgcc aaaaatctt	2520
gagggtcttg gacgcgacga gaaacgcgcg gcaaaccgcgc tgcaacaaga aaacagcgaa	2580
aatgaaagtc actgtggagt gctgggtggaa cttgagggtg acaacgcgcg cctagccgtg	2640
ctgaaacgcgca gcatcgaggt cacccactt gcctaccggg cacttaacct accccccaag	2700
gttatgagca cagtcatgag cgagctgatc gtgcgcgcgtg cacgaccctt ggagaggat	2760
gcaaacttgc aagaacaaac cgaggaggcc ctacccgcag ttggcgatga gcagctggcg	2820
cgctggcttg agacgcgcga gcctgcgcgc ttggaggagc gacgcgaagct aatgatggcc	2880
gcagtgcttg ttaccgtgga gcttgcgtgc atgcagcggg tctttgcgtga cccggagatg	2940
cagcgcaagc tagaggaaac gttgcactac acctttcgcc agggctacgt gcgccaggcc	3000
tgcaaaattt ccaacgtgga gctctgcaac ctggcttcctt accttggaaat tttgcacgaa	3060
aaccgcctcg ggcaaaacgt gcttcattcc acgctcaagg gcgaggcgcc cccgcactac	3120
gtccgcgact gcgttactt atttctgtgc tacacctggc aaacggccat gggcgtgtgg	3180
cagcaatgcc tggaggagcg caacctaaag gagctgcaga agctgtaaa gcaaaacttg	3240
aaggacctat ggacggccctt caacgagcgc tccgtggccg cgcacctggc ggacattatc	3300
ttcccccgaac gcctgtttaa aaccctgcaaa cagggtctgc cagacttac cagtcgaagc	3360
atgttgcaaa acttttaggaa ctttatccta gagcgttcaag gaattctgcc cgccacatgc	3420
tgtgcgttccctc tagcgactt tggcccatt aagtaccgtg aatgccttcc gcccgttgg	3480
ggtcactgtt accttctgca gctagccaaat taccttgcctt accactccga catcatggaa	3540
gacgtgagcg gtgacggcctt actggagtgt cactgtcgct gcaacccatg ccccgccac	3600
cgcccttgcgttgcattt gcaactgtttt agcgaaagtc aaattatccg taccttttag	3660
ctgcagggtc cctcgctga cgaaaatgc gcccgtccggg gggtgaaact cactccgggg	3720
ctgtggacgt cggcttaccc tcgcaaaatgtt gtacctgagg actaccacgc ccacgagatt	3780
aggttctacg aagaccaatc ccgcgcgcgc aatgcggagg ttaccgcctg cgttattacc	3840
caggggccaca tccttggcca attgcaagecc atcaacaaag cccgccaaga gtttctgcta	3900
cggaaaggac ggggggttta cctggacccc cagtcggccg aggagctcaa cccaatcccc	3960
ccgcgcgcgc agccctatca gcagccgcgg gcccgttgcctt cccaggatgg caccaaaaaa	4020
gaagctgcag ctgcgcgcgc cgcacccac ggacgaggagg gaataactggg acagtcaaggc	4080
agaggagggtt ttggacgagg aggaggagat gatgaaagac tggacagcc tagacgaaac	4140
ttcccgaggcc caagaggtgtt cagacgaaac accgttaccc tcggtegtcat tccccctcgcc	4200
ggcgccccag aaattggcaaa ccgttcccgat catcgctaca acctccgcgc ctcaggcgcc	4260
gccggcactg cctgttcgc gacccaaaccg tagatggac accactggaa ccagggccgg	4320
taagtctaaag cagccgcgcgc cgttagccca agagcaacaa cagcgcacaag gctaccgcgc	4380
gtggcgcggg cacaagaacg ccatagttgc ttgcttgcac gactgtgggg gcaacatctc	4440
tttcgcggccgc cgctttcttc tctaccatca cggcggtggcc ttcccccgtatc acatcctgca	4500

-continued

ttactaccgt catctctaca gcccctactg caccggcggc agcggcagcg gcagcaacag 4560
 cagcggtaac acagaagcaa aggccgaccgg atagcaagac tctgacaaag cccaaagaat 4620
 ccacagcggc ggcagcagca ggaggaggag cgctgcgtct ggcccccac gaacctgtat 4680
 cgaccggcga gcttagaaat aggattttc ccactctgtat tgctataattt caacaagca 4740
 ggggccaaga acaagagctg aaaataaaaaa acaggtctct ggcgtccctc accccgagct 4800
 gcctgtatca caaaagcgaa gatcagcttc ggccacgcg ggaagacgcg gaggctct 4860
 tcagcaaata ctgcgcgtg actcttaagg actagttcg cgcctttct caaatthaag 4920
 cgccaaaact acgtcatctc cagccggccac accccggcgcg agcacctgtc gtcagccca 4980
 ttatgagcaa ggaaattccc acgccttaca tgtggagttt ccagccacaa atgggactt 5040
 cggctggagc tgcccaagac tactcaaccc gaataaaacta catgagcgcg ggacccccaca 5100
 tgatatcccg ggtcaacggaa atccgcgcaccc accgaaacccg aattctccctc gaacaggcgg 5160
 ctattaccac cacacccgt aataaccccttata atcccccgtat ttggcccgct gcccgggtgt 5220
 accagggaaag tcccgatccc accactgtgg tacttccctc agacgcgcag gccaaggatcc 5280
 agatgactaa ctcaggggcg cagttgcgg cggcccttcg tcacagggtg cggccgcgg 5340
 ggccgttttag ggccggagtaa cttgcatgtt ttggaaattt tagttttttt aaaaatggaa 5400
 gtgacgtatc gtggggaaac ggaagtgaag atttgaggaa gttgtgggtt ttttggcttt 5460
 cgtttctggg cgttaggttcg cgtccgggtt tctgggtgtt ttttggac ttttggcttt 5520
 acgtcattttt ttagtccat atatactcgc tctgtacttg gccccttttta cactgtgact 5580
 gattgagctg gtgcgtgtc gagttgtgtt ttttaatagg tttttttact ggtaaggctg 5640
 actgttatgg ctgcgcgtgt ggaagcgcgt tatgtgttc tggagcggga ggggtctatt 5700
 ttgccttaggc aggagggttt ttcaggtgtt tatgtgtttt tcttccatataatgggttt 5760
 atacctcta tggggcgtgt aatgtgtct ctacgcgtc gggatgttat tccccggcc 5820
 tatttcggtc gtttttttagc actgaccgtt gtttaccaac ctgtatgtttt taccgagtt 5880
 tacattatgtt ctccggacat gacccggaa ctgtccgggtt tgctttttaa tcacgggtac 5940
 cagttttttt acggtaacgc cggcatggcc gtagtccgtt ttatgtttt aagggttgtt 6000
 ttccctgttg taagacaggc ttctaatgtt taaatgtttt tttttttttt atttttttt 6060
 gtgtttatgtt caggaacccg cagacatgtt tgagagaaaa atgggtgtttt tttctgtgtt 6120
 gttccggaa cttacccgtt tttatctgtca tgagcatgtac tacatgtgtc ttgtttttt 6180
 ggcgcaggct ttgcgttatttttggccgat caccgttgcattttatccgc cggccatgc 6240
 acaagcttac ataggggctt cgcgtggat ctagctccg agtatgcgtg tcataatcg 6300
 tgggggttcttttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6360
 gcacgattat gttcagctgg ccctgcgttccggat ggcgttgcatttttgcattttatccgc cggccatgc 6420
 tgggggttcttttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6480
 cgcgttgcatttttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6540
 aatattccggg atttgttttttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6600
 atgggttgcatttttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6660
 gtccacttgg acgttgcgttccggat ggcgttgcattttatccgc cggccatgc 6720
 gccattatcttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6780
 ttaatagatc ttcatttttgcatttttggccgat caccgttgcattttatccgc cggccatgc 6840

US 12,383,587 B2

169**170**

-continued

gttcttccag ctcttcccgc tcctcccgta tggactcgc agaacgaatg tgtaggttg	6900
ctgggtgtgg cttattctgc ggtgggtggat gttatcaggc cagcggcgca tgaaggagtt	6960
tacatagaac ccgaagccag gggcgccctg gatgcttgc gagagtggat atactacaac	7020
tactacacag agcgagctaa gcgacgagac cggagacgcgcatctgttg tcacgcccgc	7080
acctgggttt gtttcaggaa atatgactac gtccggcggtt ccatttgca tgacactacg	7140
accaacacga tctcggttgtt ctccggcgac tcctgtacatggatcgatccgc tacctccctt	7200
tgagacagag acccgcgcta ccatactggc ggatcatccg ctgctgccccg aatgtAACAC	7260
tttgacaatg cacaacgtga gttacgtcg aggtcttccc tgcagtgtgg gatttacgct	7320
gattcaggaa tgggtgttc cctggatata ggttctgacg cgggaggagc ttgtatccct	7380
gaggaagtgt atgcacgtgt gcctgtgttg tgccaaacattt gatatcatgc cgagcatgt	7440
gatccatggt tacgagtccct gggctctcca ctgtcattgt tccagtcggc gtccctgc	7500
gtgcatacgcc ggcgggcagg ttttggccag ctgggtttagg atgggtggatg atggcgccat	7560
gtttaatcg agggttatata ggtaccggga ggtggtaatg tacaacatgc caaaaagggt	7620
aatgtttatg tccagcgtgt ttatgagggg tccggactta atctacatgc gcttgcgtt	7680
tgatggccac gtgggttctg tggccccgc catgagctt ggatacagcg ccttgcactg	7740
tgggattttg aacaatattt tgggtctgtg ctgcagttac tggctgtttaa taagtggat	7800
cagggtgcgc tgctgtgccc ggaggacaag ggcgtctcatg ctgcggggcg tgcgaatcat	7860
cgctgaggag accactgcca tgggttattt ctgcaggacg gaggccggc ggcagcgtt	7920
tattcgcgcg ctgctgcagc accaccgcgc tatcctgtatg cacgattatg actctacccc	7980
catgttaggcg tggacttccc ctgcggccgc cgttggcaaa ccgcggatgg gacagcagcc	8040
tgtggcttag cagctggaca ggcacatgaa cttaaaggcgatg ctggccgggg agtttattaa	8100
tatcactgtatg gaggcgttgg ctgcacagga aaccgtgtgg aatataacac ctaagaatata	8160
gtctgttacc catgtatgtatg tgctttttaa ggccagccgg ggagaaaggatg ctgtgtactc	8220
tgtgtgttgg gaggggaggatg gcaagggttggaa tactagggtt ctgtgagttt gattaaggta	8280
cgggtatcaa tataagctat gtgggtgtgg ggctataacta ctgaatgaaa aatgacttgc	8340
aattttctgc aattgaaaaaa taaacacgtt gaaacataac atgcaacagg ttcacgattc	8400
tttattcctg ggcaatgttag gagaaggatg aagagttggt agcaaaaaggatg tcgtgtgt	8460
atttccact ttcccaggac catgtaaaag acatagagta agtgcattacc tcgtctgtt	8520
ctgtggattc actagaatcg atgtggatg ttgccttc tgacgggtt ggagaagggg	8580
agggtgcctt gcatgtctgc cgctgtcttt gctttgcgg ctgtgtggaa gggggccgc	8640
tctggccgcg caccggatgc atctggggaaa agcaaaaaaggatg gggctcgatcc ctgtttccgg	8700
aggaatttgc aaggcggttc ttgcgtacg gggaggcaaa ccccccgttc cccgactccg	8760
ggccggccgcg gactcgaccg gggggccctg cgactcaacc ctggaaaat aaccctccgg	8820
ctacaggaggatg cggccactt aatgttttcg ctgcggccatc taaccgttca cggccgcgc	8880
ggccaggatgc caaaaaaggatg agcgcacgcg cccggcgccgc tggaaaggatg ccaaaaaggatg	8940
cgctcccccgtt tggtctgacg tcgcacaccc tgggttgcaca cggccggcggtt aaccgttgc	9000
atcacggccggat cggccggatcc cccgggttcg aaccgggttc gtcggccatg atacccttc	9060
gaattttatcc accagaccac ggaagagtgc cccgttacatg gtcgttgcgtt tgacgggtct	9120
agagcgtcaa cggccgcgc cccgttacatg gcaaggatgc cccgaccatg gagcactttt	9180
tgccgctgcg caacatctgg aaccggcgcc ggcactttcc gggccctcc accaccggcc	9240

-continued

ccggcatcac ctggatgtcc aggtacatct acggattacg tcgacgttta aaccatatga 9300
 tcagctcaact caaaggcggt aatacggtta tccacagaat caggggataa cgccaggaaag 9360
 aacatgttag caaaaggcca gcaaaaggcc aggaaccgtaa aaaaggccgc gttgtggcg 9420
 tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcaagg 9480
 tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcg 9540
 cgctcttcgt ttccgacctt gcccgttacc ggataccctgt ccgccttct cccttcgg 9600
 agcggtggcg 9660
 tccaaagctgg gctgtgtgca cgaacccccc gttcagcccc accgctgcgc cttatccgg 9720
 aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact 9780
 ggtaaacaggaa tttagcagagc gaggtatgtt ggcgggtgcta cagagtctt gaagtgg 9840
 cctaactacg gctacactag aagaacagta tttggtatct ggcgtctgct gaagccagtt 9900
 acottcggaa aaagagttgg tagcttttga tccggcaaaac aaaccacccgc tggtagccgg 9960
 ggtttttttg ttgcagca gcaaggattacg cgccggaaaaa aaggatctca agaagatct 10020
 ttgtatctttt ctacggggtc tgacgctcgtt gggacggaa actcacaatggc agggatttt 10080
 gtcatgagat tatcaaaaag gatcttcacc tagatctttt taaattaaaa atgaagtttt 10140
 aaatcaatct aaagtatata tgagtaact tggctctgaca gttaccaatg cttaatcagt 10200
 gaggcaccta tctcagcgat ctgttattt cttttttttttt tagttgcctg actccccgtc 10260
 gtgttagataa ctacgatacg ggaggcgatccatctggcc ccagtgcgtgc aatgataccg 10320
 cgagacccac gtcacccggc tccagatata tcagcaataa accagccagc cggaaggggcc 10380
 gagcgcagaa gtggctctgc aactttatcc gcctccatcc agtctattaa ttgttgcgg 10440
 gaagctagag taagtagttc gccagttaat agtttgcgc acgttgcgttgc cattgctaca 10500
 ggcacatcggtt tgacgctcgtt atggcttcat tcagctccgg ttcccaacgaa 10560
 tcaaggcgag ttacatgatc cccatgttg tgcaaaaaag cggtagctc ctccggctt 10620
 ccgatcggtt tcagaagtaa gttggccgcgtt gttttatcacttcatggttatcc ggcacactg 10680
 cataattctc ttactgtcat gccatccgtaa agatgtttt ctgtgactgg tgagttactca 10740
 accaagtcat tctgagaata gtgtatgcgg cgaccgagt gctttgcggcc ggcgtcaata 10800
 cgggataata ccgcgcacaca tagcagaact taaaagtgc tcataatgggaaaacgttct 10860
 tcggggcgaa aactctcaag gatcttacgg ctgttgcgtt ccagttcgat gtaacccact 10920
 cgtgcaccca actgatcttc agcatctttt acttccacca gctttctgg gtgagcaaaa 10980
 acaggaaggc aaaatgccgc aaaaaaggaa ataaggcgca cacggaaatg ttgaataactc 11040
 atactcttcc tttttcaata ttattgttgcgattttatcgg gttttgtctt catgagcg 11100
 tacatatttttgcgaa aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttcccg 11160
 aaaatgtccac ctaaattgttgcgaa aatccgttata ttttgcgttta aatttttgtt 11220
 aaatcagctc attttttaac caataggccgaa aatccgttata aatcaaaaag 11280
 aatagacgca gatagggttgcgaa aatccgttata aatcaaaaag 11340
 acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgt 11400
 aaccatcacc ctaatcaatgttgcgaa aatccgttata aatcaaaaag 11460
 ctaaaggggag ccccccgttgcgaa aatccgttata aatcaaaaag 11520
 aagggaaagaa agcgaaagga gcccgcgttgcgaa aatccgttata aatcaaaaag 11580

-continued

 gcgttaaccac cacacccggc ggcgttaatg cgccgctaca gggcgcgatg gatcc 11635

<210> SEQ_ID NO 46
 <211> LENGTH: 5336
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 46

ggtacccaac tccatgttta acagtccccaa	60
acagctctac agcttcctgg agcgcactc	120
taggagcgcc actttttttt gtcacttgaa aaacatgtaa	180
ctttcaataa aggcaaatgt tttttttgtt acactctcggt	240
cttgcgtct gcccgttta aaaatcaaag gggttctgcc	300
ggcagggaca cgttgcgata ctggtgttta gtgtccact	360
cggggcagct cggtaagtt ttcaactccac	420
aggtcggggc cggatattttt gaagtcgcag ttggggcctc	480
cgatacacag ggttgagca ctggAACACT atcagcgccg	540
acgtctttgtt cggagatcag atccgcgtcc	600
gtcaactttt gtagctgcct tccaaaaaag ggtgcattcc	660
caccgtatgt gcatcagaag gtgaccgtgc	720
atgaaaaggct tgatctgctt aaaagccacc	780
ccgcaagact tgccggaaaaa ctgattggcc	840
gctcggtgt tggagatctg caccacattt	900
ttgtcttagact gtccttcag cgcgcgtgc	960
acgtgtctt tatttatcat aatgtccccgt	1020
gcccgggggtt ggccggccaa cgcgcagcc	1080
gaaacgact gcaaggatcgc ctgcaggaat	1140
ctgggtgaagg tcagctgcaat cccgggtgc	1200
gccagagctt ccacttggc	1260
tggtaacttgc ccatcaacgc	1320
ggcaggctca gccccgttat caccgtgc	1380
tcctcttgcg tccgcatacc	1440
cgttacccctt ctttgcgtgc	1500
agccacat cttcttttc ttctcgatgc	1560
tcgggcttgg gagagggggc	1620
gagggtcgatg gcccgggctt	1680
tcgtcttgcg actcgagacg	1740
ggcgacggcg acggggacga	1800
ccgcgtcg ggggggttgc	1860
aggcagaaaa agatcatggaa	1920
ttcgccacca ccgcctccac	1980

-continued

cccccgctt	aggaggagga	agtgattatc	gagcaggacc	caggtttgt	aagcgaagac	2040
gacgaggatc	gctcgttacc	aacagaggat	aaaaagcaag	accaggacga	cgcagaggca	2100
aacgaggaac	aagtccggcg	gggggaccaa	aggcatggcg	actacctaga	tgtggagac	2160
gacgtgtgt	tgaagcatct	gcagcgccag	tgcgcattt	tctgcgacgc	gttcaagag	2220
cgcagcgatg	tgcccccgc	catacgccat	gtcagcctt	cctacgaacg	ccacctgttc	2280
tcaccgegcg	taccccccac	acgccaagaa	aacggcacat	gcegaceccaa	cccgegcctc	2340
aacttctacc	ccgtatTTG	cgtgccagag	gtgcttgcac	cctatcacat	cttttccaa	2400
aactgcaaga	tacccctatc	ctgcccgtgc	aaccgcacgg	gagcggacaa	gcagctggcc	2460
ttgcggcagg	gcgctgtcat	acctgtatc	gcctcgctcg	acgaagtgc	aaaaatctt	2520
gagggtctt	gacgcgacga	gaaacgcgcg	gcaaacgcctc	tgcaacaaga	aaacagcgaa	2580
aatgaaaatgc	actgtggagt	gctgggtggaa	cttgagggtt	acaacgcgcg	cctagccgt	2640
ctgaaacgca	gcatecgaggt	cacccactt	gcctacccgg	cacttaacct	accccccac	2700
gttatgagca	cagtcgtat	cgagctgtat	gtgcgcctgt	cacgacccct	ggagagggat	2760
gcaaaactgc	aagaacaaac	cgaggaggcc	ctacccgcac	ttggcgatga	gcagctggcg	2820
cgctggctt	gacgcgcgca	gcctgcgcac	ttggaggaggc	gacgcgaagct	aatgtggcc	2880
gcagtgctt	ttaccgttgc	gcttgagtgc	atgcagcggt	tctttgttgc	cccgagatg	2940
cagcgcaagc	tagaggaaac	gttgcactac	acctttgcac	agggctacgt	gcgcaggcc	3000
tgcaaaatcc	ccaacgttgc	gctctgcac	ctggctccat	accttggaaat	tttgcacgaa	3060
aaccgcctcg	ggcaaaacgt	gcttcattcc	acgctcaagg	gcgaggcgcc	ccgcgactac	3120
ttcccgcgact	gcgtttactt	atttctgtgc	tacacctggc	aaacggccat	gggcgtgtgg	3180
cagcaatgcc	tggaggaggcg	caacctaag	gagctgcaga	agctgtaaa	gcaaaacttg	3240
aaggacctt	ggacggccctt	caacgagcgc	tccgtggccg	cgcacctggc	ggacattatc	3300
ttcccggaac	gcctgtttaa	aaccctgca	cagggtctgc	cagacttac	cagtcaaaac	3360
atgttgcata	acttttaggaa	ctttatccat	gagcttcaag	gaattctgc	cgccacccgt	3420
tgtgcgttcc	ctagcgactt	tgtgcccatt	aagtaccgtt	aatgccttcc	gccgccttgg	3480
ggtcactgt	accttctgc	gctagccaa	taccttgcct	accactccg	catcatggaa	3540
gacgtgagcg	gtgacggccct	actggagtgt	cactgtcgat	gcaacccat	cacccgcac	3600
cgccccctgg	tctgcatttc	gcaactgttt	agcgaaagtc	aaattatccg	taccttgcgt	3660
ctgcagggtc	cctcgcttgc	cgaaaagtcc	gcggctccgg	ggttggaaact	cactccgggg	3720
ctgtggacgt	cggcttaccc	tgcatttttt	gtacctgagg	actaccacgc	ccacgagatt	3780
aggttctacg	aagccaaatc	ccgcggccca	aatgcggaggc	ttaccgcctg	cgttattacc	3840
caggggccaca	tccttggcca	atttgcac	atcaacaaag	ccgcaccaaga	gtttctgtca	3900
cgaaaggagc	gggggggttta	cctggacccc	cgtccggcg	aggagctcaa	cccaatcccc	3960
ccggccgcgc	agccctatca	gcagccgcgg	gcccttgcct	cccaggatgg	cacccaaaaa	4020
gaagctgcag	ctgcccgcgc	cgccacccac	ggacgaggag	gaatactggg	acagtcaggc	4080
agaggaggtt	ttggacgagg	aggaggagat	gatggaaagac	tggacagcc	tagacgaagc	4140
ttcccgaggcc	gaagaggtgt	cagacgaaac	accgttaccc	tccgtcgat	tccctcgcc	4200
ggggccccag	aaattggcaa	ccgttccat	catcgctaca	acctccgc	ctcaggcc	4260
gccggcactg	cctgttgc	gacccaaac	tagatggac	accactggaa	ccagggccgg	4320
taagtctaa	cagccgcgc	cgttagccca	agagcaacaa	cagcgccaa	gctaccgc	4380

-continued

gtggcgccggg cacaagaacg ccatacggtgc ttgtttgcaaa gactgtgggg gcaacatctc	4440
cttcgcggccgc cgctttcttc tctaccatca cggcggtggcc ttcccccgta acatcctgca	4500
ttactaccgt catctctaca gccctactg caccggcgcc agcggcagcg gcagcaacag	4560
cagcggtcac acagaagcaa aggcgaccgg atagcaagac tctgacaaaag cccaagaaaat	4620
ccacagcgcc ggcagcagca ggaggaggag cgctgcgtct ggcgcccac gaacccgtat	4680
cgaccgcga gcttagaaat aggatttttc ccactctgtat tgctatattt caacaagca	4740
ggggccaaga acaagagctg aaaataaaaaa acaggtctct ggcgccttc acccgagct	4800
gcctgtatca caaaagcgaa gatcagcttc ggccgcacgc ggaagacgcg gaggctct	4860
tcaagcaaata ctgcgcgcgt actcttaagg actagttcg cgccctttct caaatthaag	4920
cgcgaaaact acgtcatctc cageggccac accggcgcc accgcacgtc gtcagcgcca	4980
ttatgagcaa gaaaaattccc acgcctaca tgtggagtta ccagccacaa atgggacttg	5040
cggctggagc tgcccaagac tactcaaccc gaataaaacta catgagcgcg ggacccacaa	5100
tgatatcccg ggtcaacggg atccgcgcac accgaaaccg aattctctc gaacaggcgg	5160
ctattaccac cacacctcgta aataaccta atccccgtat ttggcccgct gcccgttgt	5220
accaggaaag tccccgtccc accactgtgg tacttcccg agacgcocag gccgaagtcc	5280
agatgactaa ctcagggcg cagttgcgg gcggcttcg tcacagggtg cggcgt	5336

<210> SEQ ID NO 47
 <211> LENGTH: 3201
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 47

cccgccgtt ttagggcgga gtaacttgca tgtattggga attgttagtt ttttaaaatg	60
ggaagtgacg tatacggtggaa aaacggaaatg gaagatttga ggaagtgtg ggttttttgg	120
cttcgcgttcc tggcgtaggtt ttcgcgtcg gttttctggg tgttttttgt ggactttaac	180
cgttacgtca ttttttagtc ctatataac tgcgtctgtat cttggccctt tttacactgt	240
gactgattga gctggcgccg tgtcgagtgg tgttttttaa taggtttttt tactggtaag	300
gctgactgtt atggctgccg ctgtggaaacg gctgtatgtt gttctggagc gggagggtgc	360
tatTTTgcctt aggcaggagg gtttttcagg tgtttatgtt tttttcttc ctattaattt	420
tgttataacctt cctatggggg ctgtaatgtt gtctctacgc ctgcgggtat gtattcccc	480
gggcgtatTTTt ggtcgatTTT tagactgac cgatgttaac caacctgtat tgTTTaccga	540
gtcttacattt atgactccgg acatgaccga ggaactgtcg gtgggtgttt ttaatcacgg	600
tgaccaggTTTttt ttttacggtc acgcggcat ggccgtatgc cgtcttatgc ttataagggt	660
tgtttttccgtt gttgtaaatc aggcttctaa tgtttaaatg ttttttttt tgTTTatttt	720
ttttgcgtttt aatgcaggaa cccgcagaca tgtttgagag aaaaatggtg tcttttctg	780
tggtggttcc ggaacttacc tgccttatac tgcgtggac tgcgtacgtat gtgttgcctt	840
ttttgcgtcgaa ggctttgcctt gatTTTgtt gcaacgttgc gatTTTatac tcggccccc	900
tgcacaaacgc ttacataggg gctacgtgg ttagcatagc tccgagatgt cgtgtcataaa	960
tcagtggtggg ttctttgtc atggttccctg gccccggaaatg ggccgcgcgtg gtccgtcag	1020
acctgcacga ttatgttcag ctggccctgc gaagggaccc acgggatcgc ggtatTTT	1080

US 12,383,587 B2

179**180**

-continued

ttaatgttcc	gctttgaat	cttatacagg	tctgtgagga	acctgaattt	ttgcaatcat	1140
gattcgctgc	ttgaggctga	aggtggaggg	cgcctctggag	cagatttta	caatggccgg	1200
acttaatatt	cgggatttgc	ttagagacat	attgataagg	tggcgagatg	aaaattattt	1260
gggcatgggt	gaaggtgctg	gaatgtttat	agaggagatt	caccctgaag	ggtttagcct	1320
ttacgtccac	ttggacgtga	gggcagtttgc	cctttggaa	gccattgtgc	aacatcttac	1380
aaatgccatt	atctgttctt	tggctgtaga	gtttgaccac	gccacccggag	gggagcgcgt	1440
tcacttaata	gatcttcatt	ttgaggtttt	ggataatctt	ttggaataaa	aaaaaaaaaa	1500
catggttctt	ccagctcttc	ccgctcttc	cgtgtgtgac	tcgcagaacg	aatgtgttagg	1560
ttggctgggt	gtggcttatt	ctgccccgtt	ggatgttatac	agggcagcgg	cgcacatgaagg	1620
agtttacata	gaacccgaag	ccaggggggcg	cctggatgt	ttgagagagt	ggatatacta	1680
caactactac	acagagcggag	ctaagcgacg	agacccggaga	cgcagatctg	tttgtcacgc	1740
ccgcacccgg	ttttgttca	ggaaaatatga	ctacgtccgg	cgttccat	ggcatgacac	1800
tacgaccaac	acgatctcg	ttgtotcggc	gcactccgt	cagtagggat	cgcctaccc	1860
cttttgagac	agagacccgc	gctaccatac	tggaggatca	tccgctgtg	cccgaatgt	1920
acactttgac	aatgcacaac	gtgagttacg	tgcgagggtct	tccctgoagt	gtgggattta	1980
cgcgtattca	ggaatgggtt	gttccctggg	atatggttct	gacgcggggag	gagcttgtaa	2040
tcctgaggaa	gtgtatgcac	gtgtgcctgt	gttgcaccaa	cattgatatac	atgaegagca	2100
tgatgatcca	tggttacgag	tcctgggctc	tccactgtca	ttgttccagt	cccggttccc	2160
tgcagtgcac	agccggccgg	cagggtttgg	ccagctgggt	taggatgggt	gtggatggcg	2220
ccatgtttaa	tcaaggggtt	atatggtacc	gggaggtgg	gaattacaac	atgcacaaag	2280
aggtaatgtt	tatgtccacgc	gtgttatgt	ggggtcgcac	cttaatctac	ctgcgctgt	2340
ggtatgtgg	ccacgtgggt	tctgtggcc	ccgcacat	ctttggat	agcgccttgc	2400
actgtggat	tttgaacaat	attgtgggtc	tgtgctgcag	ttactgtgt	gatttaagt	2460
agatcagggt	gogctgtgt	gcceggagga	caaggcgct	catgctgcgg	gcgggtcgaa	2520
tcatcgctga	ggagaccact	gccatgtgt	attcctgcag	gacggagcgg	cgccggcagc	2580
agtttattcg	cgcgcgtctg	cagcaccac	gccttaccc	gatgcacat	tatgactcta	2640
cccccatgt	ggcgtggact	tcccttcgc	cgcccggt	gcaacgcac	gttggacac	2700
agccgtggc	tcaagcgact	gacagcgaca	tgaacttaag	cgagctgccc	ggggagtt	2760
ttaatatcac	tcatgtggcgt	ttggctcgac	aggaaaccgt	gtgaaatata	acacctaaga	2820
atatgtctgt	tacccatgtat	atgtgcttt	ttaaggccac	ccggggagaa	aggactgtgt	2880
actctgtgt	ttggggggg	gggtggcaggt	tgaataactag	ggttctgt	gtttgat	2940
ggtacgggt	tcaatataag	ctatgtgggt	gtggggctat	actactgaat	aaaaaatgac	3000
ttgaaatttt	ctgcaattga	aaaataaaca	cgttggaaaca	taacatgca	caggttcac	3060
attctttatt	cctggcaat	gtaggagaag	gtgttaagagt	tggttagcaaa	agtttcagtg	3120
gtgtat	tttccacca	ggaccatgt	aaagacatag	agtaagtgt	tacctcgct	3180
gtttctgtgg	attcactaga	a				3201

<210> SEQ ID NO 48

<211> LENGTH: 743

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 48

tcgatgtagg atgttgc	cccc tcctgacg	cg gtaggaga	aag gggagg	gtc cctgc	atgtc	60
tgccgctgt	cgttgc ttt	cg cgtgt	ga ggaggggg	gc atctg	ccg cagccg	120
tgc atctggg	aaa agcaaaa	aaggggctcg	tccctgtt	tc cgaggaa	t tgcaagc	180
gtcttgc	atg acggggaggc	aa acccccgt	tcgccc	cgact ccggcc	cgagactc	240
accgggggtc	ctgcgactca	acccttggaa	aataaccctc	cggtacagg	gagc gagcca	300
cttaatgtt	tcgcttcca	gcctaaccgc	ttacgccc	cgccg	ccagt gg	360
gttagcgc	ag cagccg	ccgc	gccttggaa	agg aagccaa	agc gctccc	420
acgtcgcaca	cctgggttcg	acacgcgggc	ggtaaccgca	tggatcacgg	cgga cggccg	480
gatccgggt	tgcgaa	cccc	gtcg	tccgc	atgatacc	540
cacggaa	agag	tgccgc	tta cagg	tgc	atgagcgt	600
gcacgc	cctca	ccggcc	agag	cgtcc	atggagc	660
tggaa	ccgc	tccgc	gactt	tccgc	gccttgc	720
tccagg	taca	tctacgg	attt	acg		743

<210> SEQ ID NO 49

<211> LENGTH: 160

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 49

aaggggctcg	tccctgtt	tc cgaggaa	attt tgcaagc	ggg gt	gtcttgc	atg acggggaggc	60
aa acccccgt	tcgccc	cgact	ccgc	ccggcc	cgagactc	accgggggtc	120
acccttggaa	aataaccctc	cggtacagg	gagc gagcca				160

<210> SEQ ID NO 50

<211> LENGTH: 162

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 50

aaaggagcgc	tccccgtt	tctgacgtcg	cacac	ctggg	ttcgacacgc	ggcggtaac	60
cgc	atggatc	acggcg	gacg	ccggat	ccgg	ttcgatcg	120
cccttgcgaa	tttatccacc	agaccacg	gaa	agatgt	cccg	ct	162

<210> SEQ ID NO 51

<211> LENGTH: 1590

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 51

ttaaaaatca	aagggttct	gccgcgc	atc	gtatgc	cc	actggcagg	acacgttgc	60
atactgggt	gttgt	gttcc	actt	aaactc	agg	cacaacc	atccgcggca	120

-continued

gttttcaactc cacaggctgc gcaccatcac caacgcgttt agcaggtcgg gcgccgatat	180
cttgaagtgc cagttggggc ctccgcctg cgcgcgcgag ttgcgataca cagggttgc	240
gcacttggAAC actatcagcg ccgggtggtg cacgctggcc agcacgtct tgcggagat	300
cagatccgcg tccaggctt ccgcgttgc cagggcgaac ggagtcaact ttggtagctg	360
ccttccaaa aagggtgcatt gccaggctt tgagttgcac tcgcacccgtatggcatcag	420
aagggtgaccg tgcccggtct gggcggttagg atacagcgcc tgcataaag ccttgcgt	480
cttaaaagcc acctgaggct ttgcgcctt agagaagaac atgcgcgaag acttgcggga	540
aaactgtattt gccggacagg ccgcgtcatg cacgcgcac cttgcgtcg tggtagat	600
ctgcaccaca ttgcggcccc accgggttctt cacgatcttgc cttgtctactgtcctt	660
cagcgcgcgc tgcccggttt cgctcgatca atccattca atcactgtct ctttatttat	720
cataatgcgc cctgttagac acttaagctc gccttcgatc tcagcgcagc ggtgcagcca	780
caacgcgcag ccgtgggtct cgtgggtgtt gttaggttacc tgcataaagc actgcaggta	840
cgcgtgcagg aatgcgcggca tcatcgatca aaaggcttgc ttgcgtggta aggtcagctg	900
caacccgcgg tgctccctgt ttgcgcggat cttgcatacgc ggcgcgcagg cttccacttgc	960
gtcaggcagt agcttgcattt ttgcgttttag atcggttaccc acgtggtaact tgcgttcaaa	1020
cgcgcgcgcgc gcctccatgc ctttccca cgcagacacgc atcggcaggc tcagegggtt	1080
tatcaccgtt ctttactttt ccgttcaactt ggacttccctt ttttccctt ggcgtccgcatt	1140
accccgcgc acgtgggtgtt ctttccatcg cccgcgcacc gtgcgttac cttccgttgc	1200
gtgcgttgcattt agcaccgggtt gggttgcattt acccaccatt tgcgttgcac catttctt	1260
ttttccctcg ctgtccacga tcacccctgg ggatggcgccgc ggcgtgggtt tggggaggg	1320
ggcgttcttt ttctttttgg acgcaatggc caaatccgc gtcgagggtcg atggccgcgg	1380
gttgggtgtt cgcggcaccgc ggcgttgcattt tgacgagttt ttttccctt cggactcgag	1440
acgcgcgcgc acgcgtttt ttgggggcgc gggggggaggc ggcggcgcacgc ggcacgggg	1500
cgacacgtcc tccatgggtt gtggacgtcg cgcgcgcacc gtcgcgcgtt cgggggttgtt	1560
ttcgcgtgc tccctttccc gactggccat	1590

<210> SEQ ID NO 52
 <211> LENGTH: 2446
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 52

tcccttctcc tataggcaga aaaagatcat ggagtcgttc gagaaggagg acagccatcac	60
cgcgcgcgcgc gaggatgcgc ccacccgcctc caccatgcgc gccaacgcgc ctaccaccc	120
cccccgtcgag gcaccccccgc ttgaggaggaa ggaagtgtt atcgagcagg acccagggtt	180
tgtaaaggaa gacgacgagg atcgctcgtt accaacagag gataaaaagc aagaccaggaa	240
cgacgcagag gcaaacgcagg aacaagtcgg gcggggggac caaaggcatg ggcactaccc	300
agatgtggga gacgacgtgc tggtaagca tctgcagcgc cagtgccca ttatctgcga	360
cgcgttgcgc gaggcgcaggcg atgtggccct cgcgcgcgc gatgtcaggcc ttgcgcgttgc	420
acgcgcgcgc ttttcaccgc gcttacccca caaacgcgc gaaaacggca catgcgaggcc	480
caacccgcgc ctcaacttctt accccgttattt tgcgtgcgc gagggtgttgc ccacccatca	540

-continued

catcttttc caaaaactgca agataccct atcctgccc gccaaccgca gccgagcgga	600
caaggcagctg gccttgcggc aggccgctgt cataacctgat atcgccctgc tcgacgaagt	660
gccccaaatc tttgagggtc ttggacgcga cgagaaacgc gcccggaaacg ctctgcaaca	720
agaaaaacagc gaaaatgaaa gtcactgtgg agtgctgggta gaacttgagg gtgacaacgc	780
gccccttagcc gtgctgaaac gcagcatcga ggccacccac tttgcctacc cggcacttaa	840
cctacccccc aagggttatga gcacagtcat gagcggctgtg atcgtgcgcc gtgcacgacc	900
cctggagagg gatgcaaaact tgcaagaaca aaccggaggag ggccctacccg cagttggcga	960
tgaggcagctg ggcgcgtggc ttggagacgcg cgaggcctgcg gacttggagg agcgcacgcaa	1020
gctaattatgtg gcccagtgcc ttgttaccgt ggagcttgag tgcatgcagg ggtttttgc	1080
tgacccggag atgcagcgca agcttagagga aacgttgcac tacaccccttgc gccagggtta	1140
cgtgcgcgcag gcctgcaaaa tttccaaacgt ggagctctgc aacctggctc cctacccctgg	1200
aattttgcac gaaaaccggcc tcggggcaaaa cgtgcttcat tccacgctca agggcgaggc	1260
gcgcgcgcgcac tacgtccgcg actgctgttta cttatccctg tgctacaccc ggccaaacggc	1320
catggggcgtg tgccagcaat gcctggagga ggcgcaccta aaggagctgc agaagctgct	1380
aaagccaaac ttgaaggacc tatggacggc cttaaacggag cgctccgtgg ccgcgcaccc	1440
ggcgccgacatt atcttccccg aacgcctgttataaaacccttgc caacagggtc tgccagactt	1500
caccagtcaa agcatgttgc aaaacttttag gaactttatc ctagagcggtt caggaattct	1560
ggccgcacc tacgtgtgcgc ttccctagcga ctttgtgcctt attaagtacc gtgaatggcc	1620
tccggccgtt tgggttactact gctaccccttgc acatcccttgc aactaccccttgc cttaccactc	1680
cgacatcatg gaagacgtga gcccgtacgg cctactggag tgcactgtgc gctgcaccc	1740
atgcaccccg caccgttccg tggcttgcaat ttgcacactg cttagegaaa gtcaaattat	1800
cggttacccctt gacgtcgagg gtccctgcgc tgacgaaaatggcc cttccggctc cgggggtgaa	1860
actcactccg gggctgtggc cgtccggctta ctttcgcacaa tttgttccctg aggactacca	1920
cgcccaacggat attaggttct acgaagacca atcccgcccc ccaaatgcgg agcttacccgc	1980
ctgcgtcatt acccaggggcc acatcccttgc ccaattgcaaa gccatcaaca aagccggcca	2040
agagtttctg ctacgaaagg gacgggggggt ttacctggac ccccaactgcg gcgaggagct	2100
caaccccaatc ccccccgcgc cgcagcccta tcagcggcccg cggggcccttg cttcccgagga	2160
tggcacccaa aaagaagctg cagctggcgc cggccgcacc cacggacggag gaggaatact	2220
gggacagtca ggcagaggag gtttggacg aggaggagga gatgatggaa gactgggaca	2280
gcctagacga agcttcccgag gcccggagg tgccagacga aacaccgtca ccctcggtcg	2340
cattcccttc gcccggccccc cagaattgg caaccgttcc cagcatcgct acaaccctccg	2400
ctcctcaggc gcccggggca ctgcctgttc gcccggccaa ccgttag	2446

<210> SEQ ID NO 53
<211> LENGTH: 375
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 53

gcccctactg caccggccgc agcggcagcg gcagcaacag cagcggtcac acagaagcaa	60
aggcgaccgg atagcaagac tctgacaaag cccaaagaaat ccacagcgcc ggcagcagca	120

-continued

ggaggaggag cgctgcgtct ggcccccac gaaccgtat cgaccgcga gcttagaaat	180
aggatttttc ccactctgta tgctatattt caacaaagca ggggccaaga acaagagctg	240
aaaataaaaa acaggtctct cgcgtccctc acccgagct gcctgtatca caaaagcgaa	300
gatcagcttc ggccacgct ggaagacgct gaggctctc tcagcaaata ctgcgcgtg	360
actcttaagg actag	375

<210> SEQ ID NO 54
<211> LENGTH: 354
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 54

at gagcaagg aaattccccac gcccatacg tggagttacc agccacaaat gggacttgcg	60
gctggagctg cccaaagacta ctcaacccga ataaaactaca tgagcgcggg accccacatg	120
at atccccggg tcaacggaat ccgcgcacccac cggaaaccgaa ttctcctcga acaggcggct	180
attaccacca cacctcgtaa taaccttaat ccccgtagtt ggcccgctgc cctgggttac	240
caggaaagtc ccgcgtccac cactgtggta cttcccagag acgcccaggc cgaagttcag	300
atgactaact cagggcgca gcttgcgggc ggcttcgtc acagggtgcg gtgc	354

<210> SEQ ID NO 55
<211> LENGTH: 1164
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 55

atgactacgt ccggcggtcc atttggcatg acactacgac caacacgatc tcgggtgtct	60
ccggcgactc cgtacagtag ggatgccta cctcctttt agacagagac ccgcgtacc	120
atactggagg atcatccgct gctgccgaa tgtaacactt tgacaatgca caacgtgagt	180
tacgtgegag gtctccctg cagtgtggta ttacgtcta ttccaggatg gggtgttccc	240
tgggatatgg ttctgacgct ggaggagctt gtaatcctga ggaagtgtat gcacgtgtc	300
ctgtgttgtc ccaacattga tatcatgac agcatgtatc tccatggta cgagtcctgg	360
gctctccact gtcattgttc cagtcgggt tccctgcagt gcatagccgg cggcaggtt	420
ttggccagct gttttaggt ggtgggtggat ggcgcctatg ttaatcagag gtttatatgg	480
taccgggagg tggtaatta caacatgca aaagaggtaa tgtttatgtc cagcgtgtt	540
atgaggggtc gcaacttaat ctacactgcg tttgtgtatg atggccacgt gggttctgt	600
gtccccggca tgagctttgg atacagcgcc ttgcactgtg ggatttgaa caatattgt	660
gtgtgtgtct gcaagttactg tgctgattta agtgagatca ggggtgcgtg ctgtgcgg	720
aggacaaggc gtctcatgtc gccccgggtc cgaatcatcg ctgaggagac cactgcgt	780
ttgttattccct gcaggacgga gccccgggtc cagcgttta ttgcgtgcgt gctgcagcac	840
caccgcctc tcctgatgca cgattatgac tctacccccc tggatggcgat gacttccct	900
tcgccccccg ttgagcaacc gcaagttgaa cagcagcctg tggctcagca gctggacagc	960
gacatgaact taagcgagct gccccgggag tttatataata tcactgtatga gcgtttggct	1020

US 12,383,587 B2

189**190**

-continued

cgacaggaaa ccgtgtggaa tataaacacct aagaatatgt ctgttaccct tgatatgatg	1080
cttttaagg ccagccgggg agaaaggact gtgtactctg tgtgttggga gggaggtggc	1140
aggttgaata ctagggttct gtga	1164

<210> SEQ ID NO 56
<211> LENGTH: 345
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 56

atggtttttc cagctttccc cgcttcccgtgtgtact cgcagaacga atgtgttagt	60
tggctgggtg tggcttattc tgcgggtgtg gatgttatca gggcaggggc gcatgaagga	120
gtttacatag aacccgaagc cagggggcgc ctggatgttt tgagagagtg gatatactac	180
aactactaca cagagcggac taagcgcacga gacccggagac gcagatctgt ttgtcacgcc	240
cgcacacctgtt tttgcttcag gaaatatgac tacgtccggc gttccatttg gcatgacact	300
acgaccaaca cgtatctcggt tgcgtcgccg cactccgtac agtag	345

<210> SEQ ID NO 57
<211> LENGTH: 319
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 57

atgatttcgt gcttgggtcttggag ggcgtctgg agcagatttt tacaatggcc	60
ggacttaata ttcgggattt gcttagagac atattgataa ggtggcgaga tgaaaatttat	120
ttgggcatgg ttgaagggtgc ttggatgttt atagaggaga ttcaccctga agggtttagc	180
ctttacgtcc acttggacgt gagggcagtt tgcctttgg aagccattgt gcaacatctt	240
acaaatgccttatctgttcc tttggctgtta gagtttgacc acgccaccgg aggggagcgc	300
gttcaactaa tagatcttc	319

<210> SEQ ID NO 58
<211> LENGTH: 411
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 58

atgcaggaac ccgcagacat gtttgagaga aaaatggtgt ctttttctgt ggtgggtccg	60
gaacttacct gcctttatct gcatgagcat gactacgtatg tgcttgcttt tttgcgcgag	120
gtttgcctg attttttagt cagcaccttg catttatat cggccggccat gcaacaagct	180
tacatagggg ctacgtgggt tagcatagct ccgagatgtgc gtgtcataat cagtgtgggt	240
tctttgtca tggttcctgg cggggaaagtgcggcgctgg tccgtgcaga cctgcacgat	300
tatgttcagc tggccctgcg aaggaccta cgggatcgccgtatgtttaatgttccg	360
cttttgaatc ttatacaggt ctgtgaggaa cctgaatttt tgcaatcatg a	411

<210> SEQ ID NO 59

-continued

```

<211> LENGTH: 7327
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 59

gcgcgcccgt atcgatccacg cccccggccg gcccgtctag aacttagtgaa tccccggaa      60
gatcagaagt tccttattccg aagttccat tctctagaaa gtataggaac ttctgtatcg      120
ccgcaggccgc atgcgggggt tttacgagat tgtgattaag gtcccccagcg accttgacga      180
gcatactgcgc ggcatttcgt acagctttgt gaactgggtg gcccggaaagg aatggggat      240
gcgcgccagat tctgacatgg atctgaatct gattgagcag gcacccctga ccgtggccga      300
gaagctgcacg cgccgactttc tgacggaatg gcccgtgtg agtaaggccc cggaggccct      360
tttctttgtg caatttgaga agggagagag ctacttccac atgcacgtgc tcgtggaaac      420
cacccgggggtt aaatccatgg ttttggacg tttcttgatg cagattcgcg aaaaactgtat      480
tcagagaatt taccggggga tcgagccac tttgccaac tgggtcgccg tcacaaagac      540
cagaaatggc gccggggggc ggaacaaggt ggtggatgag tgctacatcc ccaattactt      600
gtctcccaaa acccagccgt agctcagtg ggctgtggact aatatggaaac agtatttaag      660
cgccctgtttt aatctcacgg agcgtaaaacg gttgggtggcg cagcatctga cgcacggtgc      720
gcagacgcac gagcagaaca aagagaatca gaatcccaat tctgatgcgc cggtgatcag      780
atcaaaaact tcagccaggt acatggagct ggtgggtgg ctcgtggacca aggggattac      840
ctcggagaag cagtggatcc aggaggacca ggccctcatac atctccttca atgcggcctc      900
caactcgcgg tcccaaatac aggtgtgcctt ggacaatgcg ggaaagatgg tgagcctgac      960
taaaacccgc cccgactacc tgggtggcca gcagccgtg gaggacattt ccageaatcg      1020
gatttataaa attttggaaac taaacgggtt cgtatcccaat tatgcccgtt ccgttttctt      1080
gggatggggcc acgaaaaagt tcggcaagag gaacaccatc tggctgtttg ggctgtcaac      1140
taccggggaa accaacatcg cggaggccat agccccacact gtgccttct acgggtgcgt      1200
aaacttggacc aatgagaact ttcccttcaa cgactgtgtc gacaagatgg tcatctggtg      1260
ggaggaggggg aagatgaccg ccaagggtcggtt ggagtccggc aaagccatc tcggaggaaag      1320
caaggtgcgc gtggaccaga aatgcaagtc ctggcccaag atagacccga ctcccgat      1380
cgtcaccccttcc aacaccaaca tgggtggccgtt gattgacgggg aactcaacga ctttgcgaa      1440
ccagcagccg ttgcaagacc ggatgttcaa atttgaactc accccggccgtc tggatcatga      1500
ctttggggaaag gtcaccaagc aggaagtcaa agacttttc cgggtggccaa aggatcacgt      1560
gggtgagggtg gaggcatgaat tctacgtcaa aaagggtggaa gccaagaaaa gacccggccc      1620
cagtgcgc gatataagtg agcccaaaccg ggtgcgcgcg tcagttgcgc agccatcgac      1680
gtcagacgcgc gaaatgtcgatc tcaactacgc agacaggtaa caaaacaaat gttctcgatca      1740
cgtgggcgtt aatctgtatgc tgggtggccgtt cagacaatgc gagagaatga atcagaattc      1800
aaatatctgc ttcaactacgc gacagaaaga ctgttttagag tgctttcccg tgcataatc      1860
tcaacccgtt tctgtcgatc aaaaggcgta tcagaaactg tgctacatc atcataatcat      1920
gggaaagggtg ccacacgcgtt gcactgcgtt cgtatctgggtt aatgtggatt tggatgactg      1980
catctttgaa caataaatgaa tttaaatcg gtatggctgc cgatggttat cttccagatt      2040
ggctcgagga cactctctctt gaaggaataa gacagtgggtg gaagctcaaa cctggccac      2100

```

-continued

caccaccaaa	gcccgcagag	cggcataagg	acgacagcag	gggtcttgtg	cttcctgggt	2160
acaagtacct	cgAACCCCTC	aacggactcg	acaagggaga	gcccgtcaac	gaggcagacg	2220
cccgccgcct	cgAGCACGAC	aaAGCCTACG	accggcagct	cgacagcgga	gacaaccgt	2280
acctcaagta	caACCACGCC	gacgcggagt	tTCAGGAGCG	cTTAAAGAA	gatacgttt	2340
ttggggcaa	cTCTGGACGA	gcAGTCTTC	aggcgaaaaa	gagggttctt	gaacctctgg	2400
gcctgggtga	ggaACCTGTT	aAGACGGCTC	cgggaaaaaa	gaggcggta	gagcactctc	2460
ctgtggagcc	agACTCTCC	tcGGGAACCG	gaaaggcggg	ccAGCAGCCT	gcaagaaaaa	2520
gattgaattt	tggtcAGACT	ggAGACGCG	actcagtacc	tgACCCCCAG	cCTCTCGAC	2580
agccaccagc	agCCCCCTC	ggtctggaa	ctaatacgt	ggctacaggc	agtggcgcac	2640
caatggcaga	caATAACGAG	ggcGCCGACG	gagtggtaa	tTCCTGGGA	aattggcatt	2700
gcgatTCAC	atggatgggc	gacagagtca	tcaccaccag	cACCCGAACC	tggccctgc	2760
ccacctacaa	caACCACCTC	tacaaacaaa	tttccagcca	atcaggagcc	tcgaacgaca	2820
atcaCTACTT	tggctacAGC	accCTTGGG	ggtatTTGA	cttcaacAGA	ttCCACTGCC	2880
actTTTCACC	acgtgactgg	caaAGACTCA	tcaacaacaa	ctggggatTC	cgACCCAAGA	2940
gactcaACTT	caAGCTCTT	aacATTCAAG	tcaaAGAGGT	cacgcagaAT	gacggtaCgA	3000
cgacgattgc	caataACCTT	accAGCACGG	ttcaggtgtt	tactgactcg	gagtaccAGC	3060
tcccgtacgt	cTCTGGCTCG	gCGCATCAAG	gatgcctccc	gCCGTTCCC	gcAGACGTCT	3120
tcatgggcc	acAGTATGG	tacCTCACCC	tgaacaACG	gagtCAGGCA	gtaggacgct	3180
cttcatttta	ctgcCTGGAG	tactttcTT	ctcagatgt	gcgtaccggA	aacaacttta	3240
ccttcagcta	cactTTGAG	gacgttccTT	tccacAGCAG	ctacgctcac	agccAGAGTC	3300
tggaccgtct	catGAATCCT	ctcatcgacc	agtacCTGT	ttacttTGAGC	agaacaacaa	3360
ctccaAGTGG	aaccACCCACG	cagtcaAGGC	ttcagTTTTC	tcaggCCGGA	gcgagtgaca	3420
ttcggggacca	gtcttaggaAC	tggCTTCCTG	gaccCTGTTA	ccGCCAGCAG	cgagtatcaa	3480
agacatCTGC	ggataACAAC	aacAGTGAAT	actcgtggac	tggagCTACC	aagtaccacc	3540
tcaatGGCAG	agACTCTCTG	gtGAATCCGG	gcccggccat	ggcaAGCCAC	aaggacgtat	3600
aagaaaaAGT	tttCCTCAG	agcgggggttC	tcatCTTGG	gaAGCAAGG	tcAGAGAAAA	3660
caaATGTGG	cattGAAAAG	gtcatgatta	cagacGAAGA	ggAAATCAGG	acaACCAATC	3720
ccgtggctac	ggAGCAGTAT	ggttCTGTAT	ctaccaACCT	ccAGAGAGGC	aacAGACAAAG	3780
cagCTACCGC	agATGTCAAC	acacaAGGCG	ttCTTCCAGG	catGGTCTGG	caggACAGAG	3840
atgtgtacCT	tcAGGGGCC	atCTGGGCAA	agATTCCACA	cacGGACGGA	cattttcacc	3900
cctCTCCCT	catGGGTGG	ttcggactta	aaACCCCTC	tccACAGATT	ctcatcaaga	3960
acacCCCGT	acCTGCGAAT	cTTCGACCA	cTTCACTGC	ggCAAAGTT	gtttCTTCA	4020
tcacacAGTA	ctCCACGGGA	caggTCAGCG	tggAGATCGA	gtggggAGCTG	cagaAGGAAA	4080
acAGCAAACG	ctggAAATCCC	gaaATTCACT	acACTTCCAA	ctacaACAA	tctgttaATG	4140
tggactttac	tgtGGACACT	aatGGCGTGT	attcAGAGCC	tgcCCCCATT	ggcAccAGAT	4200
acctgactcg	taATCTGTA	ttgCTTGT	taAtaATAAC	cgtttaATC	gtttcAGTTG	4260
aactttggTC	tctGCGTATT	tctttCTT	ctagTTCCA	tggCTACGT	gataAGTAGC	4320
atggcgggtt	aatCATTAAAC	taCAGCCCGG	gcgttAAAC	agcGGGGGOGGA	ggggGTGGAGT	4380
cgtgacgtGA	attacgtcat	agggttaggg	aggTCCTGTA	ttagAGGTC	cgtgAGTGT	4440
ttgcgcacatt	ttgcgcacacc	atgtggCTC	gCTGGGGGGG	ggggccccGAG	tgagcAcGCA	4500

-continued

gggtctccat tttgaagcgg gaggttgaa cgagcgctgg cgcgctcaact ggccgtcggtt	4560
ttacaacgtc gtgactggga aaaccctggc gttacccaac ttaatcgct tcgcacat	4620
cccccttcg ccagctggcg taatagcgaa gagggccgca ccgatcgccc ttcccaacag	4680
ttgcgcagcc tgaatggcga atggaaatttga agcggttaa tattttgtta aaattcgct	4740
taaattttg ttaaatcagc tcattttttt aaccaatagg ccgaaatcgg caaaatccct	4800
tataaatcaa aagaatagac cgagataggg ttgagtgttg ttccagtttgc gaacaagagt	4860
caactattaa gaacgtggac tccaacgtca aaggcgaaa aaccgtctat cagggcgatg	4920
gcccaactacg tgaaccatca ccctaataa gtttttggg gtcgagggtgc cgtaaagcac	4980
taaatcgaa ccctaaaggag agcccccgat ttagagcttg acggggaaag ccggcgaacg	5040
ttggcgagaaa ggaagggaag aaagcgaaaag gagcgggcgcc tagggcgctg gcaagtgttag	5100
cggtcacgct gcgcgttaacc accacaccccg ccgcgtttaa tgccgcgcta cagggcgctg	5160
cagggtggcac tttcgggaa aatgtgcgcg gaacccttat ttgttttattt ttctaaatac	5220
attcaaatat gtatccgctc atgagacaat aaccctgata aatgcttcaa taatattgaa	5280
aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattccctt tttgcggcat	5340
tttgccttc tggtttgtct cacccagaaa cgctggtaa agtaaaagat gctgaagatc	5400
agttgggtgc acgagtgggt tacatcgaa tggatctcaa cagcggtaa atccttgaga	5460
gttttcgccc cgaagaacgt ttccaaatga tgagcacttt taaagttctg ctatgtggcg	5520
cggttattatc ccgtatttgc gcccggcaag agcaactcgcc tcgcccata cactatttcc	5580
agaatgactt ggtttagtac tcaccagtca cagaaaagca tcttaeggat ggcgtacag	5640
taagagaatt atgcagtgtc gccataacca tgagtgataa cactgcggcc aacttacttc	5700
tgacaacgat cggaggaccg aaggagctaa ccgtttttt gcacaacatg gggatcatg	5760
taactcgctc tgatcggtgg gaacggagc tgaatgaagc cataccaaac gacgagcgctg	5820
acaccacgat gctgttagca atggcaacaa cggtgcgcaaa actattaact ggcgtactac	5880
ttactctagc ttcccgccaa caattaatag actggatggg ggccgataaa gttgcaggac	5940
cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggcg	6000
agcgtgggtc tcgcggatc attgcagcac tggggccaga tggtaagccc tcccgatcg	6060
tagttatcta cacgaegggg agtcaggca ctatggatga acgaaataga cagatcgctg	6120
agatagggtgc ctcactgatt aacgttggt aactgtcaga ccaagtttac tcataatatac	6180
tttagattga tttaaaactt cattttaat ttaaaaggat cttagtgaag atccttttg	6240
ataatctcat gacaaaatc ccttaacgtg agtttcgtt ccactgagcg tcagaccccg	6300
tagaaaaatg caaaggatct tcttgagatc cttttttct gcgcgtatc tgctgttgc	6360
aaacaaaaaa accaccgcta ccagcggtgg ttgtttggcc ggatcaagag ctaccaactc	6420
ttttccgaa ggtaactggc ttcaagcagatc gctacatcgatc ttcttagtgc	6480
agccgttagtt agggcaccac ttcaagaact ctgttagcacc gcctacatcgatc tccgtctgc	6540
taatcctgtt accagtggct gctgcagatc gctacatcgatc ttcttagtgc	6600
caagacgata gttacggat aaggcgccggc ggtcgccggctg aacgggggggt tcgtgcacac	6660
agcccgatggt ggagcgaaacg acctacacccg aactcgatcgatc cctacacccg gagctatcgat	6720
aaaacggccac gttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg	6780
gaacaggaga ggcgtacgagg gagcttccag gggaaacgc ctggatctt tatacgatcgat	6840

US 12,383,587 B2

197

198

-continued

tcgggttcg ccacctctga cttgagcgta gattttgtg atgctcgta gggggccga	6900
gcctatggaa aaacgcgcgc aacgcggcct tttacgggt cctggcctt tgctggcctt	6960
ttgctcacat gtttttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct	7020
ttgagtgagc tgataccgct cgccgcagec gaacgaccga ggcgcaggag tcagtgcgcg	7080
aggaagcggaa agagcgcaca atacgcaaac cgccctccccc cgccgcgttgg ccgattcatt	7140
aatgcagctg gcacgcagg ttccccact ggaaageggg cagtgcgcg aacgcattaa	7200
atgtgagttt gctcaactcat taggcacccc aggcttaca ctttatgctt ccggctcgta	7260
tgttgtgtgg aattgtgagc ggataacaat ttcacacagg aaacagctat gaccatgatt	7320
acgccaa	7327

<210> SEQ_ID NO 60
<211> LENGTH: 6902
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 60	
ctagagtgcga cctgcaggca tgcaagcttg gcgttaatcat ggtcatagct gtttctgtg	60
tgaaattgtt atccgcac aattccacac aacatacggc ccggaaagcat aaagtgtaaa	120
gcctggggtg cctaattgggtt gagcttaactc acattaattt cggttgcgcgc actgcggcgt	180
ttccagtcgg gaaacctgtc gtgcgcggcgtt cattaaatggaa tcggccaaacg cgccgggaga	240
ggcggtttgc gtattggcg ctcttcgtt ccctcgctca ctgactcgct ggcgtcggtc	300
gttcggctgc ggcgcgggt atcagctcac tcaaaggcggtt taatacgggtt atccacagaa	360
tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt	420
aaaaaggccg cggtgtggc gttttccat aggctccggc cccctgacga gcatcacaaa	480
aatcgacgcgtt caagtcagag gtggcgaaac ccgcacaggac tataaagata ccaggcggtt	540
ccccctggaa gtccttcgtt ggcgttcgtt gttccgaccg tgccgcgttac cggataactg	600
tccgccttcc tcccttcggg aagegtggcg ctttctcata gtcacgcgtt taggtatctc	660
agttcggtgtt aggtcggtcg ctccaagctgtt gggtgtgtc acgaaacccccc cggttgcgcgc	720
gaccgcgtcg ccttatccgg taactatcgat cttgagtcga acccggttaag acacgactta	780
tcgcccactgg cagcagccac tggtaacagg attagcagag cgaggatgtt aggcgggtgt	840
acagagttct tgaagtgggtt gcctaactac ggctacacta gaaggacagt atttgggtatc	900
tgcgtctgc tgaagccagt taccttcggaa aaaagatgtt gtagctctgtt atccggcaaa	960
caaaccacccg ctggtagcggtt tgggtttttt gtttgcggc agcagattac ggcgcggaaa	1020
aaaggatctc aagaagatcc tttgatcttt tctacgggtt ctgacgcgtca gtggaaacgaa	1080
aactcacgtt aagggtttt ggtcatgtga ttatcaaaaa ggatcttac ctagatcctt	1140
ttaaattaaa aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac	1200
agtttagaaaa actcatcgag catcaaatga aactgcaattt tattcatatc aggattatca	1260
ataccatattttt tttggaaaaag ccgtttctgtt aatgaaggag aaaactcacc gaggcaggcc	1320
cataggatgg caagatcctg gtatcggtctt ggcattccggc ctgcgtccaaatcataataaa	1380
ccttattttt tccctcggtt aaaaataagg ttatcaagtgg agaaatcacc atgagtgcgcg	1440
actgaatccg gtgagaatgg caaaagttta tgcattttttt tccagacttg ttcaacagggc	1500

US 12,383,587 B2

199**200**

-continued

cagccattac gctcgtcatc aaaatcactc gcatcaacca aaccgttatt cattcgtgat	1560
tgcgccttag cgagacgaaa tacgcgatcg ctgttaaaag gacaattaca aacaggaatc	1620
gaatgcaacc ggcgcaggaa cactgccago gcatcaacaa tattttcacc tgaatcagga	1680
tattcttcta atacctggaa tgctgtttc ccagggatcg cagtggtgag taaccatgca	1740
tcatcaggag tacggataaa atgcttgatg gtggaaagag gcataaatc cgtagccag	1800
ttagtgtga ccatctcatac tgtaacatca ttggcaacgc taccttgcc atgtttcaga	1860
aacaactctg ggcgcatcggg ctcccatac aatcgataga ttgtcgacc tgattgccc	1920
acattatcgc gagccattt atacccatat aaatcagcat ccatgttggaa atttaatcgc	1980
ggcctagagc aagacgttcc cggttgaata tggctcatac ttttccttt tcaatattat	2040
tgaagcattt atcagggtta ttgtctcatg agcgataca tatttgaatg tatttagaaa	2100
aataaaacaaa taggggttcc ggcacattt ccccgaaaag tgccacactga cgtctaagaa	2160
accattatata catgacattt aacctataaa aataggcgta tcacgaggcc ctttcgtctc	2220
gcccgcgttc gtgtatgcgg tggaaacactc tgacacatgc agctcccgaa gacggtcaca	2280
gcttgttgt aagcggatgc cgggagcaga caagccgcgc agggcgcgc agcgggtgtt	2340
ggcgggtgtc ggggctggct taactatgcg gcatcagagc agattgtact gagagtgcac	2400
catatgcgtt gtgaaatacc gcacagatgc gtaaggagaa aataccgcatt caggcgccat	2460
tcgcccattca ggctgcgcaa ctgttggaa gggcgatcgg tgccggccctc ttgcgtattt	2520
cggcagctgg cgaaaggggg atgtgctgca aggcgattaa gttggtaac gccagggttt	2580
tcccagtcac gacgttgtaa aacgcacggcc agtgaattcg agctcggtac ccgtagccat	2640
ggaaactaga taagaaagaa atacgcagag accaaagttc aactgaaacg aattaaacgg	2700
tttatttattt aacaagcaat tacagattac gggtgaggtt acgggtgcac atggggccgg	2760
gttcagagta cacgccttct gtattaaacag caaagtccac acttgttagat ttgttagtagt	2820
tggaggtgtt ctggatctcg ggggtccago gcttgctgtt ttccctctgc agctccatt	2880
caatttccac gctgacctgt ccgggtgtt attgcgttat gaaagagttc agctttgact	2940
ggttgaaggt ggtcgagga tcccgaggta caggcgttcc ttgtatcagg atctgaggcg	3000
gaggatgttt caggccaaag ccgcctatca gcccggacgg gtggaaagttt ccgtccgtt	3060
gaggaatctt ggcccagatg ggacccctgca ggtacacgcgc ccggttctgc cagaccatac	3120
cgggtaaggc cccctggctt ttgacagttt caatttgggg agccgttcc ttgtgtgc	3180
agtttatctgc cacgataccg tattcctctg tagccacagg gtttagtgggtt ttgatttttt	3240
cctcgctgtt gagcatgaca tcgctgttat ccgcattgtc tctggcagca ttttgggttc	3300
caaaaatcag gatcccgtt ctggaaaaaa aacgctcctc gtcgttttgc ttgtgtgc	3360
tagcgatgcc aggattagcc aatgaatttcc ttccattcag atggatatttgc gtcggcagc	3420
tccaggcataa gttgttattt ttgttttgcg cgggtgttgc tgagacgcgt tggtggcggt	3480
aacagggtcc tggcagccag ttctttgcgtt gattggccat ttttttttttgc ccacccggc	3540
tgaagccccag agtctgcgtt ttgcgttgcg ctccctgtt gttgtccgcg gacaagttagt	3600
acaggtactg gtcaatcaga ggattcatca gcccggccaa gtcgttgcgt tggtgtgc	3660
tgctgtggaa aggacacgttcc tcgaagggtt aagtaaaactg gaagttgttgc cgggttctca	3720
gcatctgcga aggaaagtat tccaggcgtt agaaggagga ggcgtccacgc gctgtactac	3780
cgttgttagt gtttaggttgcg ccgtactggg gaatcatgaa cacgtccgc gggAACGGAG	3840
gcaggcagcc ctgggtggca gagccgagaa cgtacggcag ctggacttcc gatccgtaa	3900

US 12,383,587 B2

201

202

-continued

acacacctggat ggtgctggtg aggttattgg cgatggtctt ggtgccttca ttctgcgtga 3960
 cctccttgc acggatgttg aagagcttga agctgagtctt ctggggccgg aatccccagt 4020
 tgggtttat gagtcgctgc cagtcacgtg gtgaaaagtgc gcaatggaaat ctgttaaagt 4080
 caaaataccc ccagggggtg ctgtagccga agtaggttgtt gtcgttgggt gctcccccgg 4140
 atgtcccggtt ggagatttgc ttgttagaggt gggtgttgc gggtggcagg gcccagggttc 4200
 ggggtgtgtt ggtgtactt ctgtgcggca gccatgtggaa atcgcataatgc caattcccg 4260
 aggaactacc cactccgtcg ggcgccttgcgattgtctgc cattggtgccg ccaccgcctg 4320
 cagccattgtt attaggatccc acaccagagg ggcgtgtgg aggttctccg agaggttgag 4380
 ggtctggaaac tgactctgag tcgcccgtt gacccaaattt gagtctttt ctggggggct 4440
 gttggccctt ctggccatgtt cccgttaggg agtctggaga acgctgggggtt gatggctcta 4500
 ccgggtcttctt ctttccagga gccgtcttagt cgcgttccctt aaccagacccg agaggttcga 4560
 gaaccccgctt ctgggctgg aagactgtctt gcccggaggtt gcccccaaaa gacgtatctt 4620
 cttgcagacg ctccctgaaac tcggcgatcg cgtgggtata cccgcaggatc ggattgtcac 4680
 ccgcctcgacg ctgtggatcg taggcgttgc cgtgtcgag ggccgtgtcg tccgcgcgt 4740
 tgacggggctc ccccttgcgtc agtccgttgc aggggtccgg agtacttgcgtt ccaggaagca 4800
 ccagaccccg ggcgtgtcc tgctttgtt gggtggctt gggcttgggg gctccagggtt 4860
 tcagcgeccca ccactcgccga atgcctcag agaggttgctc ctcgagccaa tctggaaagat 4920
 aaccatggc agccataacctt gatttaaatc atttattgtt caaatgtgc gtcatccaaa 4980
 tccacattga ccagatcgca ggcagtgcac ggcgtctggca ctttcccat gatatgtga 5040
 atgttagcaca gtttctgata cgccttttgc acgacagaaaa cgggttgaga ttctgcacacg 5100
 ggaaagcact ctaaacagtc ttctgtccg tgagtgcacg agatattgtt attctgattc 5160
 attctctcgatc attgtctgca gggaaacagc atcagattca tgcccacgtc acgagaacat 5220
 ttgtttgtt acctgtctgc gtatgttgc gaaatgttccg cgtctgtacgt cgatggctgc 5280
 gcaactgact cgcgcaccccg tttgggtctca ctttatatctg cgtcaatgggg ggccgggtt 5340
 ttcttggctc cacccttttgc gacgtagaat tcatgttcca cctcaaccac gtgatccctt 5400
 gcccacccggaa aaaagtctt gacttccgtc ttgggtgactt tcccaaagtc atgatccaga 5460
 cggccgggtga gttcaaaattt gaacatccgg tcttgcaacg gtcgtgggtt ttcgaagggtc 5520
 gttgagttcc cgtcaatcac ggcgcacatgtt ttgggtgtgg aggtgtacat cacggggatc 5580
 ggggtctatctt gggccggagga cttgcatttc tggtccacgc gcacccgtt tccctccgaga 5640
 atggctttggcc cgcactccac gacccgttgc gtcatttcc cctccctccca ccagatcacc 5700
 atcttgcgtca cacaatcgatc gaaggaaatgg ttctcattgg tccagtttac gcacccgttag 5760
 aaggggcacag tggggctat cgcctccgcg atgtgggtt tcccggtatgt tgcaggccca 5820
 aacagccaga tgggtttccctt cttggccaaac tttttcgatgg cccatcccaag aaagacggaa 5880
 gcccgcataattt ggggatcgta cccgttttagt tccaaaattt tataaatccg attgtggaa 5940
 atgtcccttca cggggctgtc gcccaccagg tagtcgggggg cgggttttagt cagggtcata 6000
 atctttcccg cattgtccaa ggcagcccttgc atttggggacc ggcgtgtgg ggcggcattt 6060
 aaggagatgtt atgaggccgtc gtcccttgcgtt atccactgtc tctccgaggtt aatcccttg 6120
 tccacgcggcc accccgaccatgtc cttggctgaag tttttgtatctt gatcaccggc 6180
 gcatcagaat tgggattctg attcttttgc ttctgttccctt ggcgtgtgcga cacgtgcgtc 6240

-continued

agatgctgcg ccaccaaccg tttacgctcc gtgagattca aacaggcgct taaatactgt	6300
tccatattag tccacgcccc ca tggagctca ggctgggttt tggggagcaa gtaattgggg	6360
atgttagcact catccaccac cttgttcccg cctccggcgc catttcttgt ctgtgacc	6420
gccaaccagt ttggcaaagt cggtcgatc ccgcggtaaa ttctctgaat cagttttcg	6480
cgaatctgac tcaggaaacg tccaaaaccg atggattca ccccggttgt ttccacgac	6540
acgtgcgtat ggaagtagct ctctcccttc tcaaattgca caaagaaaag ggccctccgg	6600
gccttactca cacggcgcca ttccgtcaga aagtcgcgtc gcagcttc ggcacggc	6660
agggggtgcgtcaatccatcg attcagatcc atgtcagaat ctggcgccaa ctccatcc	6720
ttctcgccca cccagttcac aaagctgtca gaaatgccgg gcagatgctc gtcaaggctc	6780
ctggggacat taatcacaat ctcgtaaaac cccggcatgg cggctgcgc gatcagaagt	6840
tcctatactt tctagagaat aggaacttcg gaataggaac ttctgatctt ccggggatc	6900
ca	6902

<210> SEQ ID NO 61
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 61

ttagaccegtg ccatggctag ttgggtacca ggtcaccggc gtcgacttc cggtttcat	60
ggagaactgg tgaccggtaa ccgaccgtaa gattggaaat	100

<210> SEQ ID NO 62
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 62

tgtcaagaat acgatggaa aggactcgac aattgatggg cctgtagcgc caacgttgta	60
ctttatccaa ggtgagctca ggagtgttat cataaaatacg	100

<210> SEQ ID NO 63
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 63

tctcctgttc atccgacaaac cacactccct agttcaggga ggtcactcga gtttagagctg	60
aaaccccaact agtcacgct cgttatcaac ccggtaagt	100

<210> SEQ ID NO 64
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 64

-continued

gtggctcctt taaggcttcg cctgttcgat tgttcgctgc tgccggat cagcaaagtt 60
 gcttccgtcc taattcggcc taaggatccg ctatgcgcg 100

<210> SEQ ID NO 65
 <211> LENGTH: 100
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 65

actgcgcgttc tgtggctcct gaacaacgca aagcttcggt cgccctagtcc agtggatgg 60
 ccaaggacaa ttgcagtgtct tatctccact gctgaaacc 100

<210> SEQ ID NO 66
 <211> LENGTH: 100
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 66

gtgaaaagatg aggagtggat ttacgcactc acgaaggccc aattacccga gtatcgtaag 60
 ccaatatcgt agcccgatcc ctgtacaggg tcgcgtataa 100

<210> SEQ ID NO 67
 <211> LENGTH: 100
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 67

acgatttccg tcgttcatat gagccaccag gcttgtctct tcgcgcatt tgacacgca 60
 caccccgctgc cattccgagg tcgttagggc ttttgattt 100

<210> SEQ ID NO 68
 <211> LENGTH: 100
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 68

aacctatttg attgtacagc gtgagttacga agcgaatagg ctagtaaacac tgccctatgc 60
 aaagacgcct tggttgattc tagagctggg attggtaac 100

<210> SEQ ID NO 69
 <211> LENGTH: 100
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 69

tgaggtcatac tcctttcat ccgtgtccata acgcttcgat cttctgcag gctcccttta 60
 tggatgtttt atcacccctgt cgagcgtgca cgatggatgg 100

-continued

<210> SEQ ID NO 70
<211> LENGTH: 100
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 70

gaactccgcg tgcagcggcg agttgacacg gaaggccagg gagccaattt ccactgtct	60
gagcttagacc ccttgatagc acttgctgtg catggttgt	100

<210> SEQ ID NO 71
<211> LENGTH: 10150
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5875)..(5974)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (5875)..(5974)
<223> OTHER INFORMATION: This region may encompass 60-100 nucleotides

<400> SEQUENCE: 71

tagggaaata ggccagggtt tcaccgtaa acgccacatc ttgcgaatat atgtgttagaa	60
actgcccggaa atcgtcggt gtcaactcatgg aaaacgggtg aacaagggtg aacactatcc	120
catatcacca gtcaccgtc tttcattgcc atacggaaact ccggatgagc attcatcagg	180
cgggcaagaa tgtgaataaa ggccggataa aacttgtgtct tatttttctt tacgggtcttt	240
aaaaaggccg taatatccag ctgAACGGTC tggttatagg tacattggac aactgactga	300
aatgcctcaa aatgttcttt acgatgccat tgggatataat caacgggtgt atatccagt	360
attttttctt ccattttttt ttccctcctt agaaaaactc atcggatcata aatgaaact	420
gcaatttatt catatcagga ttatcaatac catatttttgg aaaaaggccgt ttctgtatg	480
aaggagaaaa ctcaccggagg cagttccata ggtggcaag atcctggtat cggctgcga	540
ttccgactcg tccaaacatca atacaaccta ttaatttccc ctcgtcaaaa ataagggtat	600
caagtgagaa atcaccatga gtgacgactg aatccgggtgaa atggcaaa agtttatgca	660
tttcttcca gacttgttca acaggccago cattacgctc gtcataaaaa tcactcgcat	720
caaccaaacc gttattcatt cgtgattgcg cctgagcggag gcgaaatacg cgatcgctgt	780
taaaaggaca attacaaca ggaatcgagt gcaaccggcg cagggaaact gccagcgcatt	840
caacaatatt ttcacactgaa tcaggatatt cttctaatac ctggAACGCT gttttccgg	900
ggatcgcagt ggtgagtaac catgcacatc caggagtagc gataaaatgc ttgatggtcg	960
gaagtggcat aaattccgtc agccagtttta gtctgaccat ctcatctgtt acatcatgg	1020
caacgctacc ttgcacatgt ttccagaaaca actctggcgcc atcggggttc ccataacaagc	1080
gatagattgt cgcacctgtat tgccccacat tatcgacgagc ccatttataac ccatataat	1140
cagcatccat gttggaaattt aatcgccggcc tcgacgtttc ccgttgaata tggctcattt	1200
ttttttccctt ctttaccaat gcttaatcgat tgaggcacctt atctcagcga tctgtctatt	1260
tcgttcatcc atagttgcct gactccccgt cgtgttagata actacgatac gggagggctt	1320

US 12,383,587 B2

209**210**

-continued

accatctggc cccagegctg cgatgatacc gcgagaacca cgctcaccgg ctccggattt	1380
atcagcaata aaccagccag ccggaagggc cgagcgaga agtgggtctg caactttatc	1440
cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt cgccagtaa	1500
tagttgcgc aacgttgttgc ccatcgctac aggcatcgta gtgtcacgct cgtcgttgg	1560
tatggcttca ttcaagtcgg gttcccaacg atcaaggcga gattcttcct tttcaatat	1620
tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag	1680
aaaaataaac aaataggggt cagtgttaca accaattaac caattctgaa cattatcgcg	1740
agccccattta tacctgataa tggctcataa cacccttgc ttgcctggcg gcagtagcgc	1800
ggtggtecca cctgacccca tgccgaactc agaagtgaaa cgccgtagcg ccgatggtag	1860
tgtggggact ccccatgcga gagtagggaa ctgccaggca tcaaataaaa cgaaaggctc	1920
agtcgaaaga ctgggcctt cggccggctt aattgagggg tgtcgccctt attcgactcg	1980
gggctcgagc tgcgcgctcg ctgcgtcaact gaggccgccc gggcaaagcc cgggcgtcgg	2040
gcgaccccttgc tcgcggccgc ctcagtgcgc gagcgagcgc gcagagaggg agtggccaac	2100
tccatcacta ggggttcctt taattaaacg cgtttacata acttacggta aatggccgc	2160
ctggctgacc gccccaaacgcac ccccgcccat tgacgtcaat aatgacgtat gttccatag	2220
taacgccaat agggactttc cattgacgta aatgggtggaa ctatttacgg taaactgccc	2280
acttggcagt acatcaagt tatcatatgc caagtacgcc ccctattgac gtcaatgacg	2340
gtaaatggcc cgcctggcat tatggccagt acatgacctt atgggacttt cctacttggc	2400
agtacatcta cgtattagtc atcgcttata ccatgggtat gcgggtttgg cagtagatca	2460
atgggcgtgg atagcggtt gactcacggg gatttccaag tctccacccc attgacgtca	2520
atgggagttt gttttggcac caaaaatcaac gggactttcc aaaaatgtcgta aacaactccg	2580
ccccatttgc gcaaatgggc ggtaggcggt tacgggtggga ggtctatata ggccgcgcga	2640
actgaaaaac cagaaaatgttta actggtaagt ttatgtttt tgcgttttat ttcaaggccc	2700
ggatccggtg gtgggtcaaa tcaaagaact gctcctcagt ggatgttgcc tttacttcta	2760
ggccctgtacg gaagtgttac ttctgtctta aaagctctcg cagggaattt gccaccatgg	2820
tgagcaaggcg cgaggagctg ttccacgggg tggtgcccat cttggcgag ctggacggcg	2880
acgtaaacgg ccacacaatgttc agcgtgtccg gcgaggcgca gggcgatgcc acctacggca	2940
agctgaccct gaagttcatc tgcaccaccc gcaagctgcc cgtccctgg cccaccctcg	3000
tgaccaccct gacccatggc gtgcagtgtct tcagccgctca ccccgaccac atgaagcagc	3060
acgacttctt caagtcgcgc atgcccgaag gctacgttca ggagcgaccat atcttcttca	3120
aggacgacgg caactacaag accccgcgcgc aggtgaagtt cgaggcgac accctggta	3180
accgcacatcgaa gctgaagggc atcgacttca aggaggacgg caacatccctg gggcacaagc	3240
tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaacgac aagaacggca	3300
tcaagggtcaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag ctgcggacc	3360
actaccacgca gaacacccccc atcgccgcacg gccccgtct gctgcccgcac aaccactacc	3420
tgagcaccca gtccggccctg agcaaagacc ccaacgagaa gcgcgatcac atggcctgc	3480
tggagttcgat gaccggccgc gggatcactc tcggcatggc cgagctgtac aagtaataga	3540
ctagtgcggcc ttcggccccc cccccctaa cgttactggc cgaagecgct tggataagg	3600
ccgggtgtcg tttgtctata ttttttttcc caccatattt ccgttttttgc gcaatgtgag	3660
ggcccgaaa cctggccctg tcttcttgc ggcattccctt aggggtcttt cccctctcg	3720

-continued

caaaggaatg caaggctctgt tgaatgtcgt gaaggaagca gttcctctgg aagcttcttg	3780
aagacaaaaca acgtctgttag cgaccctttg caggcagcgg aaccccccac ctggcgcacag	3840
gtgcctctgc ggccaaaagc cacgtgtata agatacacacct gcaaaggcgg cacaacccca	3900
gtgccacgtt gtgagttgga tagttgtgga aagagtcaaa tggctcacct caagcgtatt	3960
caacaagggg ctgaaggatg cccagaaggt accccattgt atgggatctg atctggggcc	4020
tcgggtcaca tgctttacat gtgttagtc gaggttaaaa aacgtctagg ccccccgAAC	4080
cacggggacg tggtttccct ttgaaaaaca cgatgataat accgggtgcca ccatgctgt	4140
gctgctgtc ctgctggcc tgaggctaca gctctccctg ggcatcatcc cagttgagga	4200
ggagaacccg gacttctgga acccgcgagcc agecgaggcc ctgggtgccc ccaagaagct	4260
gcagcctgca cagacagccg ccaagaacccat catcatcttc ctgggcatgt ggatgggggt	4320
gtctacggtg acagctgcca ggatcctaaa agggcagaag aaggacaac tggggctgaa	4380
gataccccctg gccatggacc gcttccata tggctctgtt tccaaagacat acaatgtaga	4440
caaacatgtg ccagacagtg gagccacago cacggctac ctgtgcgggg tcaaggccaa	4500
cttccagacc attggcttga gtgcagccgc ccgccttaac cagtgcacaca cgacacgcgg	4560
caacgaggc atctccgtga tgaatcgggc caagaaagca gggaaagtcaag tgggagtgg	4620
aaccaccaca cgagtgccgc acgectcgcg ageccggacc tacgcccaca cggtaaccg	4680
caactggta tcggacgccc acgtgcctgc ctggcccg caggaggggt gccaggacat	4740
cgctacgcg ctcatctca acatggacat tgacgtgatc cttagtgagg gccgaaagta	4800
catgttgcg atggaaaccc cagaccctga gtacccagat gactacagcc aaggtgggac	4860
caggctggac gggaaagaaatc tggtgccgat atggctggc aagcggccagg gtgcccggta	4920
tgtgtgaaac cgactgagc tcatgcaggg ttccctggac ccgtctgtga cccatctcat	4980
gggtctcttt gagcctggag acatgaaata cgagatccac cgagactcca cactggaccc	5040
ctccctgatg gagatgacag aggctgcct gcgcctgtcg agcaggaaacc cccggggctt	5100
cttccttcgatc gtggagggtg gtcgcacatcgatccat catgaaagca gggcttaccg	5160
ggcactgact gagacgatca tggctgacga cgccattgag agggcgggccc agtcaccag	5220
cgaggaggac acgctgagcc tcgtcaactgc cgaccactcc cacgtttct cttcggagg	5280
ctacccctcg ctagggagct ccatttcgg gctggccctt ggcaaggccc gggacaggaa	5340
ggcctacacg gtcctcttat acggaaacgg tccaggatcat gtgctcaagg acggccccc	5400
gccggatgtt accgagagcg agagcgggag ccccgagat cggcagcagt cagcagtgc	5460
cctggacgaa gagacccacg caggcggagga cgtggccgtt ttcgcgcgcg gcccgcaggc	5520
gcacctgggtt cacggcgtgc aggaggacat cttcatcgatcg caccgtcatgg cttcgcgcgc	5580
ctgcctggag ccctacacccg cctgcgcaccc ggcgcggggcc gcccgcgc	5640
gcacccgggt tactctagag tggggccggc cggccgccttc gagcagacat gagtcgacag	5700
atctttaaa aaacctccca cacaattgtt gttgttaact tggatttgc agcttataat	5760
ggttacaat aaagcaatag catcacaaat ttccacaataa aagcattttt ttcaactgc	5820
tctagttgtt gttgtccaa actcatcaat gtatcttac atgtctgtt aaacnnnnnn	5880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn	5940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnatttaa ataggaaccc cttagtgatgg	6000
agttggccac tccctctctg cgcgctcgat cgctcaactga ggccggccga ccaaaggctg	6060

-continued

cccgacgccc	gggctttgcc	cggggggcct	cagtgagega	gcgagcgccg	agagagtata	6120
catcgatgtg	agttcgcggg	tggctgggg	gccccggcct	gcaaccgcg	6180	
tctacgagcc	ttgcgggctc	cgggtcttgc	cagtcgtatg	ggggcagggt	agctgttccc	6240
cgcaaggaga	gctcaagggtc	agcgcgtcgg	cctggcggag	ccccgcaccc	aggctgtggc	6300
gccccgtgtca	gatccggccct	tgcggcgcca	tctgccccgg	gcttccttcc	cctagtcccc	6360
agaaacagga	ggtccctact	cccgcggag	atcccgaccc	ggacccctag	gtgggggacg	6420
ctttcttcc	tttcgegctc	tgcgggggtca	cgtgtcgac	aggagccct	ccccacggc	6480
ctccggcacc	gcaggccccg	ggatgctagt	gcgcagcggg	tgcateccctg	tccggatgct	6540
gcccctgcgg	tagagcggcc	gccatgttgc	aaccgggaag	gaaatgaatg	ggcagccgtt	6600
aggaaaggct	gcccgtact	aacctgcgc	tctgcctcg	atgggtggag	tgcgtgtgg	6660
cggggaagtc	aggtggagcg	aggctagctg	gcccgatttc	tctccgggt	gatgttttc	6720
ctagattatt	ctctggtaaa	tcaaagaagt	gggttatgg	aggtccttctt	gtgtcccctc	6780
cccgcagagg	tgtgggtggct	gtggcatgtt	gccaagccgg	gagaagctga	gtcatggta	6840
gttggaaaag	gacatttcca	ccgcaaaatg	gccccctctgg	tggtgccccc	ttccctgcagc	6900
gcccggctcac	ctcacggccc	cgccccctccc	ctgccagccct	agcgttgacc	cgaccccaaa	6960
ggccaggctg	taaatgtcac	cgggaggatt	gggtgtctgg	gcgcctcg	gaacctgccc	7020
ttctccccc	tccgtttcc	ggaaaccaga	tctcccaccg	caccctggc	tgagggttaaa	7080
tatagctgt	gacccctctg	tagctgggg	cctggctgg	ggctctctcc	catccctct	7140
ccccacacac	atgcacttac	ctgtgtcc	actccctatt	tctggaaaag	agcttaggaag	7200
gacaggcaac	ttggcaaatac	aaagccctgg	gactaggggg	ttaaaataca	gctccctc	7260
ttccccacccg	cccccagcttc	tgtccctttt	gtaggaggg	cttagagaag	gggtggcctt	7320
gcccctgtcca	gttaatttct	gacccctact	cctgccctt	gagtttgatg	atgctgagtg	7380
tacaaggcgtt	ttctccctaa	agggtgcago	tgagcttaggc	agcagcaagc	attccctggg	7440
tggcatagtg	gggtggtgaa	taccatgtac	aaagcttgc	cccagactgt	gggtggcagt	7500
cccccacatg	gcccgttctc	ctggaaggcc	tgcgtatgac	tgggggtgtt	gggcagccct	7560
ggagccttca	gttgcagcca	tgccttaagc	caggccagcc	tggcaggaa	gctcaaggaa	7620
gataaaaatc	aaccccttgg	gccctctgg	gggtaaggag	atgctgcatt	cgcccttta	7680
atggggaggt	ggccttagggc	tgctcacata	ttctggagga	gctccctc	ctcatgcct	7740
cttgcctctt	gtctcttagg	catgcaaaag	agtcgaataa	ggcgacaca	aaatttattc	7800
taaatgcata	ataaaatactg	ataacatctt	atagttgtt	ttatattttt	tattatcg	7860
gacatgtata	attttgcata	aaaaactga	tttccctt	attattttcg	agatttattt	7920
tcttaattct	ctttaacaaa	ctagaaatat	tgtatataca	aaaaatcata	aataatagat	7980
gaatagttt	attatagg	ttcatcaatc	aaaaagcaa	cgatctttat	ttaaaagtgcg	8040
ttgtttttt	ctcatttata	aggtaataata	attctcatat	atcaagcaa	gtgacaggcg	8100
cccttaataa	ttctgacaaa	tgctcttcc	ctaaactccc	cccataaaaa	aacccgcga	8160
agcgggtttt	tacgttattt	gccccgat	gattactcg	tatcagaacc	gcccagggg	8220
cccgagctt	agactggccg	tcgttttaca	acacagaaag	agtttgtaga	aacgcaaaa	8280
ggccatcggt	caggggcctt	ctgcttagtt	tgcgtatgg	cagttcccta	ctctcgccct	8340
ccgcttctc	gtcactgac	tgcgtgcgt	cggtcgatcg	gctgcggcga	gcccgtatcg	8400
ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	ggataacgc	ggaaagaaca	8460

-continued

tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcgcccc	8520
tccatagcgct	ccggccccc	gacgagcata	acaaaaatcg	acgctcaagt	cagagggtggc	8580
gaaaacccgac	aggactataa	agataccagg	cgttcccccc	tggaagctcc	ctcgctcgct	8640
ctccctgtcc	gaccctgccc	cttaccggat	acctgtccgc	ctttctccct	tcggaaagcg	8700
tggcgcttcc	tcatagctca	cgctgttaggt	atctcagttc	ggtgttaggtc	gttcgctcca	8760
agctgggctg	tgtgcacgaa	ccccccgttc	agcccgacc	ctgcgcctta	tccggtaact	8820
atcgctttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	8880
acaggattag	cagagcgagg	tatgttaggcg	gtgctacaga	gttcttgaag	tgggtggcta	8940
actacggcta	cactagaaga	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	9000
tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaac	caccgcgttgt	agcgggtggtt	9060
tttttggtttgc	caagcagcag	attacgcgca	gaaaaaaaaagg	atctcaagaa	gatccttga	9120
tcttttctac	gggggtctgac	gctcagtgg	acgacgcgc	cgtaactcac	gttaaggaggat	9180
tttggtcatg	agcttgcgc	gtcccgtaa	gtcagcgtaa	tgctacac	tttcccgat	9240
gccccacttt	gtatagagag	ccactgcggg	atcgtcaccc	taatctgc	gcacgttagat	9300
ttataaagca	ccaagcgcgt	ttggcctttag	ctttagggaga	ttgatgagcg	cggtggcaat	9360
gccccctgc	cggtgcgc	cgaggactgc	gagatthaag	atatacgat	cactacgcgg	9420
ctgctcaaac	ttggcagaa	cgtaagccgc	gagagcgc	acaacccgtt	cttggtcgaa	9480
ggcagcaagc	gcgtatgt	tcttacta	gagcaagttc	ccgaggtaat	cgaggatccgg	9540
ctgtatgtgg	gagtaggtgg	ctacgtcacc	gaactcacg	ccgaaaagat	caagagcagc	9600
ccgttagat	ttgacttgg	cagggccgag	cctattatgt	cgatgtatgc	cttaacttga	9660
gccaccta	tttggtttag	ggcgactgc	ctgctgcgt	attacgttgc	tgctttaaa	9720
ttacaaacat	cgaccacgg	cgtaacgcgc	ttgctgcgttgc	gtgcccgg	gcatacgact	9780
tacaaaaaaa	cagtcataac	aagccatgaa	aaccgcact	gcccgttac	caccgcgtgc	9840
ttcggtaa	gttctggacc	agttgcgtga	gcgcattttt	ttttctcct	cgcgcttac	9900
gccccgc	gcactcatac	gcagtaactgt	tgtaattcat	taagcattct	gccgacatgg	9960
aaggccatc	agacggcatg	atgaaacctg	atgcgcagcg	gcatcagcac	cttgcgcct	10020
tgcgtataat	atttgc	atttgc	agtaaaaacg	ggggcgaaga	agttgtccat	10080
tttaaatcaa	aactggtgaa	actcacccag	ggattggcgc	tgacgaaaaa	catattctca	10140
ataaaccctt						10150

<210> SEQ ID NO 72
<211> LENGTH: 498
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 72

cacttctcc	cgtatgccc	actttgtata	gagagccact	gcgggatcgt	caccgtatc	60
tgcttgcacg	tagatttata	aagcaccaag	cgcgttggcc	tttagcttga	ggagattgtat	120
gagcgcgg	gcaatgc	gcctccgg	ctcgccgg	actgcgagat	ttaagatata	180
gatctca	cgcgg	cttgcgt	caaacttgg	cagaacgtaa	gccgcgagag	240
cgcttctgg	tcgaaggcag	caagcgc	aatgtctta	ctacggagca	agttcccgag	300

-continued

gtaatcggag tccggctgat gttgggagta ggtggctacg tcaccgaact cacgaccgaa	360
aagatcaaga gcagccccgtt aggatttgac ttggtcaggg ccgagccatat tagtgcgaat	420
gatgccttaa cttgagccac ctaactttgt ttttagggcga ctgccctgct gcgtaattac	480
gttgctgctc ttaaattta	498
<210> SEQ ID NO 73	
<211> LENGTH: 7989	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide	
<220> FEATURE:	
<221> NAME/KEY: modified_base	
<222> LOCATION: (3714)...(3813)	
<223> OTHER INFORMATION: a, c, t, g, unknown or other	
<220> FEATURE:	
<221> NAME/KEY: misc_feature	
<222> LOCATION: (3714)...(3813)	
<223> OTHER INFORMATION: This region may encompass 60-100 nucleotides	
<400> SEQUENCE: 73	
tagggaaata ggccagggtt tcaccgtaac acgccacatc ttgcgaatat atgtgttagaa	60
actgcccggaa atcgtcgtgt gcactcatgg aaaacgggtt aacaagggtt aacactatcc	120
cataatcacca gtcaccggtc tttcattgcc atacggaact ccggatgagc attcatcagg	180
cgggcaagaa tgtgaataaa ggccggataa aacttgtgtt tattttttt tacggcttt	240
aaaaaggccg taatatccag ctgaacgggtc tggttatagg tacattgagc aactgactga	300
aatgcctcaa aatgttcttt acgatgccat tgggatataat caacgggttgt atatccagt	360
atttttttctt ccattttttt ttcctccctt agaaaaactc atcgagcatc aaatgaaact	420
gcaatttttt catatcagga ttatcaatac catattttg aaaaaggccgt ttctgtat	480
aaggagaaaa ctcaccgagg cagttccataa ggatggcaag atcctggat cggctgcga	540
ttccgactcg tccaacatca atacaaccta ttaatttccc ctgcgtaaaa ataaggttat	600
caagttagaa atcaccatga gtgacgactg aatccgggtga gaatggcaaa agtttatgca	660
tttcttcca gacttgttca acaggccago cattacgctc gtcatcaaaa tcactcgcat	720
caaccaaacc gttattcatt cgtgttgcgc cctgagcggag cgaaataacg cgatcgctgt	780
taaaaggaca attacaaaca ggaatcgagt gcaaccggcg caggaacact gccagcgcat	840
caacaatatt ttcacctgaa tcaggatatt cttctaatac ctggAACGCT gttttccgg	900
ggatcgcagt ggtgagtaac catgcatcat caggagtacg gataaaatgc ttgatggtcg	960
gaagtggcat aaatccgtc agccgatTTA gtctgaccat ctcatctgta acatcatgg	1020
caacgcgtacc tttgccatgt ttcagaaaca actctggcgc atcggggttc ccataacaagc	1080
gatagattgt cgcacctgat tgcccgacat tatcgccgac ccatttataac ccataaaat	1140
cagcatccat gttggaaattt aatcgccggcc tcgacgttcc cggttgaata tggctcattt	1200
tttttccctc ctttaccaat gcttaatcg tgaggccacct atctcagcga tctgtctatt	1260
tcgttcatcc atagttgcct gactccccgt cgtgttagata actacgataac gggagggctt	1320
accatctggc cccagcgctg cgatgataacc gcgagaacca cgctcacccg ctccggattt	1380
atcagcaata aaccagccag ccggaaaggcc cgagcgcaga agtggctctg caactttatc	1440
cgcctccatc cagtcttattt attgttgcgc ggaagctaga gtaagtagtt cgccagtaa	1500
tagtttgcgc aacgttggcc ccatcgctac aggcacgtg gtgtcacgct cgctcggttgg	1560

US 12,383,587 B2

219**220**

-continued

tatggcttca ttcagctccg gttcccaacg atcaaggcga gattcttcct tttcaatat	1620
tattgaaca ttatcaggg ttattgtctc atgagcggat acatatttga atgtatTTAG	1680
aaaaataaaac aaataggggt cagtgttaca accaattaac caattctgaa cattatcg	1740
agcccattta tacctgaata tggctcataa cacccttgc ttgcctggcg gcagtagc	1800
ggtgttcca cctgaccca tgccgaactc agaagtgaaa cgccgtagcg ccgatggtag	1860
tgtggggact ccccatgcga gagtagggaa ctgccaggca tcaaataaaa cgaaaggctc	1920
agtcgaaaga ctgggcctt cggccgggt aattgagggg tgtcgccctt attcgactcg	1980
gggctcgagc acgagctgcg cgctcgctcg ctcaactgagg ccggccgggc aaagccccgg	2040
cgtcggcgcga ctttggctcg cccggctca gtgagcgcgc gagegagcag agagggagtg	2100
gggttttaat taaacgcgtt tacataactt acggtaatgg gcccgcctgg ctgaccggcc	2160
aacgaccccc gcccattgac gtcaataatg acgtatgttc ccatagtaac gccaataggg	2220
actttccatt gacgtcaatg ggtggactat ttacggtaaa ctgcccactt ggcagtacat	2280
caagtgtatc atatgcgaag tacggccctt attgacgtca atgacggtaa atggcccgcc	2340
tggcattatg cccagttacat gaccttatgg gacttcctca ctggcagta catctacgt	2400
ttagtcatcg ctattaccat ggtgatgcgg tttggcagt acatcaatgg gcgtggatag	2460
cgggtttact cacggggatt tccaaagtctc caccctattt acgtcaatgg gagtttttt	2520
tggcaccaaa atcaacggga ctttccaaaaa tgctgttaca actccgcoccc attgacgaa	2580
atgggcggta ggcgtgtacg gtggaggta tatataggcg cgccgaactg aaaaaccaga	2640
aagtttaactg gtaagtttag tctttttgtc ttttatttca ggtccggat ccgggtgg	2700
tgcaaatcaa agaactgctc ctcagtggtat gttgccttta cttctaggcc tgcgtggaa	2760
tgttacttct gctctaaaag ctctgcagg gaattgcaca ccatggtgag caagggcgag	2820
gagctgttca cgggggtggt gccatcctg gtctcgatgg acggcgcacgt aaacggccac	2880
aagttcagcg tgcgtggcgaa gggcgaggcc gatgccacct acggcaagct gaccctgaag	2940
ttcatctgca ccacccggaa gctgcggctg ccctggccca ccctcgatgc caccctgacc	3000
tacggcgtgc agtgcttcag ccgttacccc gaccatgtg agcagcacga cttttcaag	3060
tccggcatgc ccgaaggcta cgccaggag cgccaccatct tcttcaagga cgacggcaac	3120
tacaagaccc ggcggaggt gaagttcgag ggccacacc tggtgaaccg catcgatcg	3180
aaggccatcg acttcaagga ggacggcaac atccctgggg acaagctgga gtacaactac	3240
aacagccaca acgtctatcat gatggccgac aagcagaaga acggcatcaa ggtgaacttc	3300
aagatccgccc acaacatcgaa ggacggcago gtgcagctcg ccgaccacta ccagcagaac	3360
accccccattcg ggcgacggccc cgtgtgtcg cccgacaacc actacctgag caccctgacc	3420
ccctgtggca aagaccccaa cgagaagcgc gatcacatgg tcctgtggta gttcgatgg	3480
ggccggccggta tcaacttcgg catggacgag ctgtacaagt aatagactag tgcgtggaa	3540
tcttttaaaa aacctccac acaattgttg ttgttaactt gtttatttca gcttataatg	3600
gttacaaata aagcaatagc atcacaaattt tcacaaataa agcattttt tcaactgtcatt	3660
ctagttgtgg tttgtccaaa ctcatcaatg tatcttatac tgcgtgttta aacnnnnnnn	3720
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn	3780
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnatttaaa taggaacccc tagtgtatgg	3840
gttggccact ccctctctgc ggcgtcgctc gctcaactgag gccggccgac caaaggctcg	3900

US 12,383,587 B2

221

222

-continued

ccgacgcccc ggcttgccc ggggggcctc agtgagcgag cgagcgcgca gagagtatac	3960
atcgatgtga gttegegggt ggctgggggg ccctgggctc cgaccgcgg cgaaccgcgt	4020
ctacgagcct tgcgggctcc gggctttgc agtcgtatgg gggcagggtt gctgtcccc	4080
gcaaggagag ctcaagggtca ggcgtcgac ctggcggagc cccgcaccca ggctgtggcg	4140
ccctgtgcag ctccgeccctt gggggccat ctggccggag cctccctccc ctatccccca	4200
gaaaacaggag gtccctactc ccgcgggaga tccccacccc gacccttagg tgggggacgc	4260
tttcttctt ttcgegctc gcgggggtcac gtgtcgcaga ggagccctc ccccacggcc	4320
tccggcaccg caggccccgg gatgtctgg cgcagcggtt gcatccctgt ccggatgctg	4380
cgcctgcgggtt agagcgccgg ccatgttgca accgggaagg aaatgaatgg gcaagccgtta	4440
ggaaaggctc cccgtgacta accctgcgtt cctgcctcga tgggtggagt cgcgtgtggc	4500
ggggaaagtca ggtggagcga ggctagctgg cccgattctt cctccgggtt atgctttcc	4560
tagattatttc tctggtaaat caaagaagtg ggtttatggg ggtcctctt tgccccctcc	4620
ccgcagagggtt gtgggtggctg tggcatggtg ccaagccggg agaagctgag tcatgggttag	4680
ttggaaaagg acatttccac cgaaaaatgg cccctctggt ggtggccctt tcctgcagcg	4740
ccggctcacc tcacggccccc gccttcccc tggcagccta gcgttgaccc gacccaaag	4800
gccaggctgtt aaatgtcacc gggaggattt ggtgtctggg cgcctcgaaaa aacctgcctt	4860
tctccccattt ccgtcttcccg gaaaccagat ctcccacccgc accctgggtt gaggttaaat	4920
atagctgctt acccttctgtt agctgggggg ctgggtctggg gctctctccc atcccttctc	4980
cccacacaca tgcacttacc tgcacttccca ctccctgattt ctggaaaaga gcttaggaagg	5040
acaggcaact tggcaaatac aagccctggg actagggggg taaaatacag cttccctct	5100
tcccacccgc cccagttctctt gtcccttttg taggaggggc ttagagaagg ggtgggttt	5160
ccctgtccag ttaatttctg acccttactc ctggccctttt agtttgcattt tgctgagttt	5220
acaagcgttt tctccctaaa ggggtcagctt gagctaggca gcagcaagca ttccctgggtt	5280
ggcatagttt ggtggtaat accatgtaca aagcttgcgc ccagactgtt ggtggcagtg	5340
ccccacatgg cgccttctcc tggaaaggctt tcgtatgactt ggggggtttt ggcagccctg	5400
gaggcttcag ttgcagccat gccttaagcc agggcagecctt ggcaggaaag ctcaaggag	5460
ataaaattca accttctggg cccttctggg ggttaaggaga tgctgcattt cccttctaa	5520
tggggaggtt gccttagggctt gctcacatat tctggaggag ccccttctcc tcatgcctt	5580
ttgcctcttg tctcttaggc atgcataaaga gtcgaataaag ggccacacaa aattttttt	5640
aaatgcataa taaatactga taacatctta tagttgtat tatattttt attatcggtt	5700
acatgtataa ttttgcatac aaaaactgtt tttcccttta ttattttgcgat tttttttt	5760
cttaatttctt ttaacaaac tagaaatattt gtatatacaa aaaatctaa ataatagatg	5820
aatagtttaa ttataggtt tcatcaatcg aaaaagcaac gtatctttt taaagtgcgtt	5880
tgcttttttc tcatttataa ggttaataa ttctcatata tcaagcaag tgacaggcgc	5940
ccttaataat tctgacaaat gctttttccc taaactcccc ccataaaaaa acccgccgaa	6000
gggggtttttt acgttatttgc cggattaacg attactcgat atcagaaccc cccaggggc	6060
ccgagcttaa gactggccgtt cggtttacaa cacagaaaga gtttgcgtt gacaggcgc	6120
ccatccgttccggcctt tgccttagttt gatgcctggc agttccctac tctgccttc	6180
cgcttcctcg ctcaactgact cgctgcgtt ggtcggttgcg ctgcggcggag cggtatcagc	6240
tcactcaaag gcgtaatac ggttatccac agaatcagg gataacgcag gaaagaacat	6300

-continued

gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gcccgcgttc tggcgtttt 6360
 ccataggctc cgccccctg acgagccatca caaaaatcgta cgctcaagtc agagggtggcg 6420
 aaacccgaca ggactataaa gataccaggo gttcccccgtt ggaagctccc tcgtgcgctc 6480
 tcctgttccg accctgcgc ttaccggata cctgtccgc ttctccctt cgggaagcgt 6540
 ggccgtttct catacgctcac gctgttaggta tctcagttcg gtgttaggtcg ttcgctccaa 6600
 gctgggtgt gtgcacgaac cccccgttca gccccgaccgc tgccgccttat ccggtaacta 6660
 tcgtctttag tccaacccgg taagacacga cttatgccta ctggcagcag ccactggtaa 6720
 caggattagc agagcggagt atgtaggcgg tgctacagag ttcttgaagt ggtggctaa 6780
 ctacggctac actagaagaa cagtatattgg tatctgcgcct ctgctgaagc cagttacctt 6840
 cggaaaaaga gttggtagct cttgtatccgg caaacaaacc accgctggta gccgggtttt 6900
 ttttgggttc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat 6960
 cttttctacg gggctctacg ctcagtgaa cgcacgcgcg ctaactcactc ttaaggatt 7020
 ttgggtcatga gcttgccgcg tcccgtaag tcagcgtaat gctacacttc ttcccgat 7080
 cccaaactttg tatagagagc cactgcggga tcgtcaccgt aatctgottt cactgttagatt 7140
 tataaaagcac caagcgcgtt ggcctttago ttgaggagat tcatgagcgc ggtggcaatg 7200
 ccctgcctcc ggtgctcgcc ggagactcgcg agatattaaga tatagatctc actacgcgc 7260
 tgctcaaact tggcagaac gtaagccgcg agagcgccaa caaccgcctt ttggtcgaag 7320
 gcagcaagcg cgtatgtt cttactacgg agcaagtcc cgaggtaatc ggagtccggc 7380
 tcatgttggg agtaggtggc tacgtcaccg aactcacgac cgaaaagatc aagagcagcc 7440
 cgtaggttggatt tgacttggtc agggccgago ctattatgcg gaatgtatgcc ttaacttgag 7500
 ccacctaact ttgttttagg gcgactgccc tgctgcgtaa ttacgttgc gctcttaat 7560
 tacaacatc gacccacggc gtaacgcgtc tgctgcgtt atgcccggg catagactgt 7620
 aaaaaaaaaac agtcataaca agccatgaaa accgcactg cgccgttacc accgcgtcgt 7680
 tcggtcaagg ttctggacca gttgcgttag cgatttttt tttccctccgc ggcgtttacg 7740
 ccccgccctg ccactcatcg cagtagtgc gtaattcatt aagcattctg ccgacatgg 7800
 agccatcaca gacggcatga tgaacctgaa tcggccagegg catcagcacc ttgtgcctt 7860
 gcgtaataata ttggccata gtgaaaacgg gggcgaagaa gttgtccata ttggccacgt 7920
 ttaaatcaaact gatggtggaaa ctcacccagg gatggcgt gacgaaaaac atattctaa 7980
 taaaccctt 7989

<210> SEQ ID NO: 74
 <211> LENGTH: 21
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
 <220> FEATURE:
 <221> NAME/KEY: MISC_FEATURE
 <222> LOCATION: (1)..(3)
 <223> OTHER INFORMATION: This region may or may not be present

<400> SEQUENCE: 74

Gly	Ser	Gly	Glu	Gly	Arg	Gly	Ser	Leu	Leu	Thr	Cys	Gly	Asp	Val	Glu	
1								5						10		15

Glu	Asn	Pro	Gly	Pro
			20	

-continued

```

<210> SEQ ID NO 75
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(3)
<223> OTHER INFORMATION: This region may or may not be present

<400> SEQUENCE: 75

```

```

Gly Ser Gly Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val
1           5           10          15

Glu Glu Asn Pro Gly Pro
20

```

```

<210> SEQ ID NO 76
<211> LENGTH: 23
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(3)
<223> OTHER INFORMATION: This region may or may not be present

<400> SEQUENCE: 76

```

```

Gly Ser Gly Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp
1           5           10          15

Val Glu Ser Asn Pro Gly Pro
20

```

```

<210> SEQ ID NO 77
<211> LENGTH: 25
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide
<220> FEATURE:
<221> NAME/KEY: MISC_FEATURE
<222> LOCATION: (1)..(3)
<223> OTHER INFORMATION: This region may or may not be present

<400> SEQUENCE: 77

```

```

Gly Ser Gly Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala
1           5           10          15

Gly Asp Val Glu Ser Asn Pro Gly Pro
20

```

What is claimed is:

1. A plasmid system for Recombinant Adeno-Associated Viral Vector (rAAV) production comprising the following three plasmids:
 - (i) a transgene-containing plasmid comprising at least one heterologous nucleic acid flanked by a 5' and 3' AAV inverted terminal repeat (ITR) and a stuffer sequence outside of the ITRs, wherein the stuffer sequence comprises a nucleic acid having at least 40% identity to SEQ ID NO: 9;

- (ii) a plasmid comprising AAV replication (Rep) and capsid (Cap) gene sequences; and
 - (iii) an adenovirus (Ad) helper plasmid, wherein the backbone of the transgene-containing plasmid is larger than a wild-type AAV genome, and wherein the transgene-containing plasmid further comprises a DNA titer tag having a length of between about 60 nucleotides to about 100 nucleotides.
2. The plasmid system of claim 1, wherein the transgene-containing plasmid is not packaged into an rAAV capsid.

227

3. The plasmid system of claim 1, wherein the stuffer sequence comprises a nucleic acid sequence of between 1000 and 2600 nucleotides in length.

4. The plasmid system of claim 1, wherein the stuffer sequence comprises GAPDH intron 2, fragment, or mutant thereof. 5

5. The plasmid system of claim 1, wherein the stuffer sequence comprises SEQ ID NO: 9.

6. The plasmid system of claim 1, wherein the transgene-containing plasmid comprises nucleic acid sequences in the 5' to 3' direction of: SEQ ID NO: 2, SEQ ID NO: 4, at least one heterologous nucleic acid, SEQ ID NO: 8, SEQ ID NO: 3, and the stuffer sequence. 10

7. The plasmid system of claim 1, wherein the transgene-containing plasmid comprises:

- a) the DNA titer tag i) upstream of the 3' ITR and downstream of a polyA sequence or ii) upstream of the 3' ITR and downstream of the at least one heterologous nucleic acid; iii) or downstream of the 5' ITR and upstream of the at least one heterologous nucleic acid sequence; or iv) downstream of the 5' ITR and upstream of a promoter for the at least one heterologous nucleic acid sequence; or v) downstream of the 5' ITR and upstream of the 3' ITR; or 15

228

b) the DNA titer tag i) upstream of SEQ ID NO: 3 and downstream of SEQ ID NO: 8 or ii) upstream of SEQ ID NO: 3 and downstream of the at least one heterologous nucleic acid sequence; iii) or downstream of SEQ ID NO: 2 or SEQ ID NO: 43 and upstream of the at least one heterologous nucleic acid sequence; or iv) downstream of SEQ ID NO: 2 and upstream of SEQ ID NO: 4; or v) downstream of SEQ ID NO: 2 or SEQ ID NO: 43 and upstream of SEQ ID NO: 3. 20

8. The plasmid system of claim 1, wherein the transgene-containing plasmid comprises nucleic acid sequences in the 5' to 3' direction of: SEQ ID NO: 43, SEQ ID NO: 4, at least one heterologous nucleic acid sequence, SEQ ID NO: 8, SEQ ID NO: 3, and the stuffer sequence.

9. The plasmid system of claim 8, wherein the transgene-containing plasmid comprises the DNA titer tag outside the expression cassette but between the 5' ITR and 3' ITR. 25

10. A host cell comprising the plasmid system of claim 1.

11. A method for producing a Recombinant Adeno-Associated Viral Vector (rAAV) comprising transducing a cell with the plasmid system of claim 1 and isolating the rAAV. 30

12. A composition comprising the plasmid system of claim 1.

* * * * *