

RETROVIRAL VECTORS

The present invention relates to retroviral gene transfer vectors, particularly lentiviral vectors, pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory
5 paramyxovirus, comprising a promoter and a transgene; and methods of making the same.

BACKGROUND TO THE INVENTION

Retroviruses are a family of RNA viruses (Retroviridae) that encode the enzyme reverse transcriptase. Lentiviruses are a genus of the Retroviridae family, and are characterised by a long
10 incubation period. Retroviruses, and lentiviruses in particular, can deliver a significant amount of viral RNA into the DNA of the host cell and have the unique ability among retroviruses of being able to infect non-dividing cells, so they are one of the most efficient methods of a gene delivery vector.

Pseudotyping is the process of producing viruses or viral vectors in combination with foreign viral envelope proteins. As such, the foreign viral envelope proteins can be used to alter host tropism
15 or an increased/decreased stability of the virus particles. For example, pseudotyping allows one to specify the character of the envelope proteins. A frequently used protein to pseudotype retroviral and lentiviral vectors is the glycoprotein G of the Vesicular stomatitis virus (VSV), short VSV-G.

Lentiviral vectors, especially those derived from HIV-1, are widely studied and frequently used vectors. The evolution of the lentiviral vectors backbone and the ability of viruses to deliver
20 recombinant DNA molecules (transgenes) into target cells have led to their use in many applications. Two possible applications of viral vectors include restoration of functional genes in genetic therapy and *in vitro* recombinant protein production.

When designing retroviral/lentiviral vectors suitable for use as gene delivery vectors, one key driver is to make the vector as safe as possible for patients. A second key driver is the need to produce
25 sufficient quantities of the vector not just to treat an individual patient, but to allow wider clinical access to the therapy for all patients who could benefit from the therapy. These two drivers can find themselves in conflict, as modifications which improve vector safety are often associated with decreased yield during vector production.

One example of a clinical setting which would benefit from gene transfer to the airway
30 epithelium is treatment of Cystic Fibrosis (CF). CF is a fatal genetic disorder caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene, which acts as a chloride channel in airway epithelial cells. CF is characterised by recurrent chest infections, increased airway secretions, and eventually respiratory failure. In the UK, the current median age at death is ~25 years. For most genotypes, there are no treatments targeting the basic defect; current treatments for symptomatic

relief require hours of self-administered therapy daily. Gene therapy, unlike small molecule drugs, is independent of CFTR mutational class and is thus applicable to all affected CF individuals. However, to date there are no viral vectors approved for clinical use in the treatment of CF, and the same applies to other diseases, particularly many other respiratory tract diseases.

5 In addition to patient safety and yield issues, there are other difficulties conventionally associated with gene transfer to the airway epithelium.

Gene transfer efficiency to the airway epithelium is generally poor, at least in part because the respective receptors for many viral vectors appear to be predominantly localised to the basolateral surface of the airway epithelium. As such, prior to the inventors' research, the use of lentiviral pseudotypes required disruption of epithelial integrity to transduce the airways, for example by the use of detergents such as lysophosphatidylcholine or ethylene glycol bis(2-aminoethyl ether)-*N,N,N',N'*-tetraacetic acid, has been linked to an increased risk of sepsis. In addition, conventional gene transfer vectors struggle to penetrate the respiratory tract mucus layer, which also reduces gene transfer efficiency. The ability to administer conventional viral vectors repeatedly, mandatory for the life-long treatment of a self-renewing epithelium, is limited, because of patients' adaptive immune responses, which prevent successful repeat administration.

Administration of the vectors for clinical application is another pertinent factor. Therefore, viral stability through use of clinically relevant devices (e.g. bronchoscope and nebuliser) must be maintained for treatment efficacy.

20 There is accordingly a need for a gene therapy vector that is able to circumvent one or more of the problems described above. In particular, it is an object of the invention to provide a method for producing a pseudotyped retroviral or lentiviral (e.g. SIV) vector, and the means for carrying out said method, wherein the resulting vector is safe and adapted for improved gene transfer efficiency across the airway epithelium, and is produced at clinically relevant scale.

25

SUMMARY OF THE INVENTION

The present inventors have previously developed a lentiviral vector, which has been pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, comprising a promoter and a transgene. Typically, the backbone of the vector is from a simian immunodeficiency virus (SIV), such as SIV1 or African green monkey SIV (SIV-AGM). Preferably the backbone of a viral vector of the invention is from SIV-AGM. The HN and F proteins function, respectively, to attach to sialic acids and mediate cell fusion for vector entry to target cells. The present inventors discovered that this specifically F/HN-pseudotyped lentiviral vector can efficiently transduce airway epithelium, resulting in transgene expression sustained for periods

beyond the proposed lifespan of airway epithelial cells. Importantly, the present inventors also found that re-administration does not result in a loss of efficacy. These features make the vectors of the present invention attractive candidates for treating diseases via their use in expressing therapeutic proteins: (i) within the cells of the respiratory tract; (ii) secreted into the lumen of the respiratory tract; and (iii) secreted into the circulatory system.

However, there were potential safety concerns with this lentiviral vector. In particular, there was a significant degree of sequence homology between the genome vector and the GagPol vector used in its production. This sequence homology creates a theoretical risk that a replication competent lentivirus (RCL) could be generated either during manufacture, or in clinical use following administration to a patient. This represents a safety risk to the patient. The risk of generating replication competent viral particles is an issue for other retroviral/lentiviral vectors as well.

Whilst it would be desirable to mitigate this risk, it is not straightforward to do so, or at least not without eliciting other unacceptable disadvantages. In particular, it is established in the art that modifications aimed at reducing the risk of RCL, such as codon-optimisation of the manufacturing gag-pol genes typically negatively impacting the titre or yield of the vector. Given the large titres of vector required to treat even a single patient, such a reduction in yield has the potential to render its production commercially unviable.

The present inventors have now demonstrated that for the first time that the use of codon-optimised gag-pol genes from SIV do not negatively impact the manufactured titre of a SIV vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, and can even result in an increased titre of the vector. This is surprising, given that under normal manufacturing conditions (when the vector genome plasmid, rather than the gag-pol genes, is limiting), codon-optimisation of the gag-pol genes typically decreases vector yield.

Therefore, the present inventors are the first to provide a method for the production of a retroviral, particularly a lentiviral vector, such as SIV, pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus with a reduced risk of RCL, without negatively affecting, or even increasing vector titre. Thus, the methods of the invention provide for safer vectors produced at commercially desirable yields.

Accordingly, the present invention provides a method of producing a retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, and which comprises a promoter and a transgene, wherein said method comprises the use of codon-optimised gag-pol genes. Preferably, the retroviral vector is a lentiviral vector, and optionally the lentiviral vector is selected from the group consisting of a Simian immunodeficiency virus (SIV) vector, a Human immunodeficiency virus (HIV) vector, a Feline immunodeficiency virus (FIV)

vector, an Equine infectious anaemia virus (EIAV) vector, and a Visna/maedi virus vector. Particularly preferred are methods of producing an SIV vector.

The codon-optimised gag-pol genes may be SIV gag-pol genes. The codon-optimised gag-pol genes may comprise or consist of a nucleic acid sequence having at least 80% sequence identity to SEQ ID NO: 1. The codon-optimised gag-pol genes may comprise or consist of the nucleic acid sequence of SEQ ID NO: 1. The codon-optimised gag-pol genes may be comprised in a plasmid that comprises or consists of a nucleic acid sequence having at least 80% sequence identity to SEQ ID NO: 5. The codon-optimised gag-pol genes may be comprised in a plasmid that comprises or consists of the nucleic acid sequence of SEQ ID NO: 5.

10 The respiratory paramyxovirus may be a Sendai virus.

The titre of retroviral vector produced by a method of the invention may be: (a) equivalent to the titre of retroviral vector produced by a corresponding method which does not use codon-optimised gag-pol genes; or (b) increased compared with the titre of retroviral vector produced by a corresponding method which does not use codon-optimised gag-pol genes. Optionally, the titre of retroviral vector may be at least 1.5-fold, at least 2-fold, or at least 2.5-fold greater than the titre of retroviral vector produced by a corresponding method which does not use codon-optimised gag-pol genes.

The promoter may be selected the group consisting of a cytomegalovirus (CMV) promoter, elongation factor 1a (EF1a) promoter, and a hybrid human CMV enhancer/EF1a (hCEF) promoter. Preferably the vector comprises a hybrid human CMV enhancer/EF1a (hCEF) promoter.

The transgene may be selected from: (a) a secreted therapeutic protein, optionally Alpha-1 Antitrypsin (A1AT), Factor VIII, Surfactant Protein B (SFTPB), Factor VII, Factor IX, Factor X, Factor XI, von Willebrand Factor, Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) and a monoclonal antibody against an infectious agent; or (b) *CFTR*, *ABCA3*, *DNAH5*, *DNAH11*, *DNAI1*, and *DNAI2*. Preferably the transgene encodes: (i) *CFTR*; (ii) *A1AT*; or (iii) *FVIII*.

In particularly preferred embodiments, the method produces a retroviral/lentiviral (e.g. SIV) vector wherein: (a) the promoter is a hCEF promoter and the transgene encodes *CFTR*; (b) the promoter is a hCEF promoter and the transgene encodes *A1AT*; or (c) the promoter is a hCEF or CMV promoter and the transgene encodes *FVIII*.

30 The method of the invention may comprise or consist of the following steps: (a) growing cells in suspension; (b) transfecting the cells with one or more plasmids; (c) adding a nuclease; (d) harvesting the lentivirus; (e) adding trypsin; and (d) purification. The one or more plasmids may comprise or consist of: (a) a vector genome plasmid, preferably selected from selected from pGM830 and pGM326 or variants thereof as defined herein; (b) a co-galpol plasmid, preferably pGM691 or

variant thereof as defined herein; (c) a Rev plasmid, preferably pGM299 or variant thereof as defined herein; (d) a fusion (F) protein plasmid, preferably pGM301 or a variant thereof as defined herein; and (e) a hemagglutinin-neuraminidase (HN) plasmid, preferably pGM303 or a variant thereof as defined herein. The ratio of vector genome plasmid: co-gagpol plasmid: Rev plasmid: F plasmid: HN plasmid may be 20:9:6:6:6.

Steps (a)-(f) of the method may be carried out sequentially. The cells may be HEK293 cells (such as HEK293F or HEK293T cells) or 293T/17 cells. The addition of the nuclease may be at the pre-harvest stage. The addition of trypsin may be at the post-harvest stage. The purification step may comprise one or more chromatography step.

The vector genome plasmid may be modified to reduce the number of retroviral ORFs.

The invention also provides a nucleic acid comprising codon-optimised gag-pol genes, said nucleic acid having at least 80% sequence identity to SEQ ID NO: 1. Preferably the nucleic acid comprises or consists of the nucleic acid sequence of SEQ ID NO: 1.

The invention further provides a plasmid comprising a nucleic acid of the invention, wherein optionally: (a) the plasmid comprises or consists of a nucleic acid sequence having at least 80% sequence identity to SEQ ID NO: 5; or (b) the plasmid comprises or consists of the nucleic acid sequence of SEQ ID NO: 5. Optionally within the plasmid the nucleic acid is operably linked to a promoter driving expression of the Gag and Pol proteins, preferably a CAG promoter.

The invention also provides a host cell comprising a nucleic acid of the invention, and/or a plasmid of the invention.

The invention further provides a retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus which is obtainable by a method of the invention.

The invention also provides a method of treating a disease comprising administering a retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus which is obtainable by a method of the invention to a subject in need thereof. The disease to be treated may be a lung disease, preferably cystic fibrosis.

BRIEF DESCRIPTION OF THE DRAWINGS

30

Figure 1: shows an alignment of the wild-type (non-codon-optimised) gag-pol genes from pGM297 with the exemplary codon-optimised gag-pol genes of the invention from pGM691, showing the changes to the wild-type sequence.

Figure 2: A-F show schematic drawings of exemplary plasmids used for production of the vectors of the invention. G shows a non-codon-optimised gag-pol plasmid (pDNA2a, specifically pGM297) that can be codon-optimised according to the invention.

- 5 **Figure 3:** shows a schematic drawings of an exemplary pDNA1 plasmid used for production of the A1AT vectors of the invention.

Figure 4: A-D show schematic drawings of exemplary pDNA1 plasmids used for production of the FVIII vectors of the invention.

10

- Figure 5:** A illustrates homology between the pDNA1 plasmid pGM326 and the non-codon-optimised pDNA2a plasmid pGM297. B compares the non-codon-optimised pDNA2a plasmid pGM297 and the codon-optimised pDNA2a plasmid pGM691 of the invention, with differences between the two annotated. C a DNA matrix homology plot illustrates homology between the DNA sequence present in pGM297 (horizontal axis) and pGM691 (vertical axis). The solid diagonal line represents sequence homology, broken line highlights areas of reduced sequence identity; note the reduced sequence identity in the areas of gag and pol gene codon optimisation in pGM691. Note also the additional sequence present in pGM297 (located approximately 6000 to 7000 bases on the numbering shown on the horizontal axis) – this is the RRE region present in pGM297 but absent in pGM691. D ClustalW DNA sequence alignment of the gag pol regions of pGM297 (lower row of DNA sequence) and pGM691 (upper row of DNA sequence); sequence homology is indicated by boxed shaded regions, a consensus DNA sequence is shown underneath the pGM691 and pGM297 sequence listings. Note the complete DNA homology between the pGM297 and pGM691 sequence in (i) the gag pol Slip region, the overlapping portion of the gag pol genes, and (ii) the rabbit beta globin poly adenylation sequence (RBG pA). Note also that pGM297 contains the SIV RRE sequence while this is absent in pGM691. E shows a restriction map of the codon-optimised gag-pol genes within the pGM693 plasmid
- 15
- 20
- 25

- Figure 6:** A shows that under design of experiment (DOE) conditions, the use of a codon-optimised pDNA2a plasmid pGM691 resulted in an observable increase in the titre of rSIV.F/HN hCEF-CFTR vector. B shows that the increase in rSIV.F/HN hCEF-CFTR vector titre obtained using the codon-optimised pDNA2a plasmid pGM691 is exhibited across two different sets of experimental conditions.
- 30

Figure 7: shows that the titre of rSIV.F/HN CMV-EGFP vector obtained using the codon-optimised pDNA2a plasmid pGM691 is greater than that obtained using the non-codon-optimised gagpol in the

pDNA2a plasmid pGM297. This suggests that the advantageous properties of codon-optimised gagpol in F/HN pseudotyped vectors is not limited to the rSIV.F/HN hCEF-CFTR, but is a general property of using codon-optimised gagpol in F/HN pseudotyped vectors.

- 5 **Figure 8:** shows a linear plasmid map for the Partial Gag RRE cPPT hCEF region of the pGM326 vector genome plasmid.

Figure 9: shows an annotated schematic of the pGM326 vector genome plasmid, with SIV ORFs identified. In particular, two large ORFs, one of 189 amino acids (aa), one of 250aa were identified
10 upstream of the hCEF promoter and soCFTR2 transgene.

Figure 10: shows that the pGM326 vector genome plasmid and modified pGM830 vector genome plasmid in otherwise identical conditions (including non-coGagPol) produce comparable vector titres in both HEK293T cells (left panel) and A549 cells (right panel).

15

Figure 11: shows the vector titre produced using coGagPol and either pGM326 or pGM830 in otherwise identical conditions, with an observable trend to increased vector titre when coGagPol is combined with pGM830.

20 **DETAILED DESCRIPTION OF THE INVENTION**

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs.
25 Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 20 ED., John Wiley and Sons, New York (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide the skilled person with a general dictionary of many of the terms used in this disclosure. The meaning and scope of the terms should be clear; however, in the event of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic
30 definition. It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such can vary.

This disclosure is not limited by the exemplary methods and materials disclosed herein, and any methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of this disclosure. The terminology used herein is for the purpose of

describing particular embodiments only, and is not intended to limit the scope of the present invention, which is defined solely by the claims.

The description of embodiments of the disclosure is not intended to be exhaustive or to limit the disclosure to the precise form disclosed. While specific embodiments of, and examples for, the disclosure are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the disclosure, as those skilled in the relevant art will recognize. For example, while method steps or functions are presented in a given order, alternative embodiments may perform functions in a different order, or functions may be performed substantially concurrently. The teachings of the disclosure provided herein can be applied to other procedures or methods as appropriate. The various embodiments described herein can be combined to provide further embodiments. Aspects of the disclosure can be modified, if necessary, to employ the compositions, functions and concepts of the above references and application to provide yet further embodiments of the disclosure. Moreover, due to biological functional equivalency considerations, some changes can be made in protein structure without affecting the biological or chemical action in kind or amount. These and other changes can be made to the disclosure in light of the detailed description. All such modifications are intended to be included within the scope of the appended claims.

Unless otherwise indicated, any nucleic acid sequences are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

The headings provided herein are not limitations of the various aspects or embodiments of this disclosure.

As used herein, the term "capable of" when used with a verb, encompasses or means the action of the corresponding verb. For example, "capable of interacting" also means interacting, "capable of cleaving" also means cleaves, "capable of binding" also means binds and "capable of specifically targeting..." also means specifically targets.

Other definitions of terms may appear throughout the specification. Before the exemplary embodiments are described in more detail, it is to be understood that this disclosure is not limited to particular embodiments described, and as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present disclosure will be defined only by the appended claims.

Numeric ranges are inclusive of the numbers defining the range. Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also

specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within this disclosure. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller
5 ranges is also encompassed within this disclosure, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in this disclosure.

As used herein, the articles "a" and "an" may refer to one or to more than one (e.g. to at least one) of the grammatical object of the article. Further, unless otherwise required by context, singular
10 terms shall include pluralities and plural terms shall include the singular. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, the use of the term "including", as well as other forms, such as "includes" and "included", is not limiting.

"About" may generally mean an acceptable degree of error for the quantity measured given the nature or precision of the measurements. Exemplary degrees of error are within 20 percent (%),
15 typically, within 10%, and more typically, within 5% of a given value or range of values. Preferably, the term "about" shall be understood herein as plus or minus ($\pm$) 5%, preferably $\pm$ 4%, $\pm$ 3%, $\pm$ 2%, $\pm$ 1%, $\pm$ 0.5%, $\pm$ 0.1%, of the numerical value of the number with which it is being used.

The term "consisting of" refers to compositions, methods, and respective components thereof as described herein, which are exclusive of any element not recited in that description of the
20 invention.

As used herein the term "consisting essentially of" refers to those elements required for a given invention. The term permits the presence of elements that do not materially affect the basic and novel or functional characteristic(s) of that invention (i.e. inactive or non-immunogenic ingredients).

25 Embodiments described herein as "comprising" one or more features may also be considered as disclosure of the corresponding embodiments "consisting of" and/or "consisting essentially of" such features.

Concentrations, amounts, volumes, percentages and other numerical values may be presented herein in a range format. It is also to be understood that such range format is used merely
30 for convenience and brevity and should be interpreted flexibly to include not only the numerical values explicitly recited as the limits of the range but also to include all the individual numerical values or sub-ranges encompassed within that range as if each numerical value and sub-range is explicitly recited.

As used herein, the terms "vector", "retroviral vector" and "retroviral F/HN vector" are used interchangeably to mean a retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, unless otherwise stated. The terms "lentiviral vector" and "lentiviral F/HN vector" are used interchangeably to mean a lentiviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, unless otherwise stated. All disclosure herein in relation to retroviral vectors of the invention applies equally and without reservation to lentiviral vectors of the invention and to SIV vectors that are pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus (also referred to herein as SIV F/HN or SIV-FHN).

As used herein, the terms "titre" and "yield" are used interchangeably to mean the amount of lentiviral (e.g. SIV) vector produced by a method of the invention. Titre is the primary benchmark characterising manufacturing efficiency, with higher titres generally indicating that more retroviral/lentiviral (e.g. SIV) vector is manufactured (e.g. using the same amount of reagents). Titre or yield may relate to the number of vector genomes that have integrated into the genome of a target cell (integration titre), which is a measure of "active" virus particles, i.e. the number of particles capable of transducing a cell. Transducing units (TU/mL also referred to as TTU/mL) is a biological readout of the number of host cells that get transduced under certain tissue culture/virus dilutions conditions, and is a measure of the number of "active" virus particles. The total number of (active+inactive) virus particles may also be determined using any appropriate means, such as by measuring either how much Gag is present in the test solution or how many copies of viral RNA are in the test solution. Assumptions are then made that a lentivirus particle contains either 2000 Gag molecules or 2 viral RNA molecules. Once total particle number and a transducing titre/TU have been measured, a particle:infectivity ratio calculated. Amino acids are referred to herein using the name of the amino acid, the three-letter abbreviation or the single letter abbreviation.

As used herein, the terms "protein" and "polypeptide" are used interchangeably herein to designate a series of amino acid residues, connected to each other by peptide bonds between the alpha-amino and carboxyl groups of adjacent residues. The terms "protein", and "polypeptide" refer to a polymer of amino acids, including modified amino acids (e.g., phosphorylated, glycosylated, etc.) and amino acid analogues, regardless of its size or function. "Protein" and "polypeptide" are often used in reference to relatively large polypeptides, whereas the term "peptide" is often used in reference to small polypeptides, but usage of these terms in the art overlaps. The terms "protein" and "polypeptide" are used interchangeably herein when referring to a gene product and fragments thereof. Thus, exemplary polypeptides or proteins include gene products, naturally

occurring proteins, homologs, orthologs, paralogs, fragments and other equivalents, variants, fragments, and analogues of the foregoing.

As used herein, the terms "polynucleotides", "nucleic acid" and "nucleic acid sequence" refers to any molecule, preferably a polymeric molecule, incorporating units of ribonucleic acid, deoxyribonucleic acid or an analogue thereof. The nucleic acid can be either single-stranded or double-stranded. A single-stranded nucleic acid can be one nucleic acid strand of a denatured double-stranded DNA. Alternatively, it can be a single-stranded nucleic acid not derived from any double-stranded DNA. In one aspect, the nucleic acid can be DNA. In another aspect, the nucleic acid can be RNA. Suitable nucleic acid molecules are DNA, including genomic DNA or cDNA. Other suitable nucleic acid molecules are RNA, including siRNA, shRNA, and antisense oligonucleotides. The terms "transgene" and "gene" are also used interchangeably and both terms encompass fragments or variants thereof encoding the target protein.

The transgenes of the present invention include nucleic acid sequences that have been removed from their naturally occurring environment, recombinant or cloned DNA isolates, and chemically synthesized analogues or analogues biologically synthesized by heterologous systems.

Minor variations in the amino acid sequences of the invention are contemplated as being encompassed by the present invention, providing that the variations in the amino acid sequence(s) maintain at least 60%, at least 70%, more preferably at least 80%, at least 85%, at least 90%, at least 95%, and most preferably at least 97% or at least 99% sequence identity to the amino acid sequence of the invention or a fragment thereof as defined anywhere herein. The term homology is used herein to mean identity. As such, the sequence of a variant or analogue sequence of an amino acid sequence of the invention may differ on the basis of substitution (typically conservative substitution) deletion or insertion. Proteins comprising such variations are referred to herein as variants.

Proteins of the invention may include variants in which amino acid residues from one species are substituted for the corresponding residue in another species, either at the conserved or non-conserved positions. Variants of protein molecules disclosed herein may be produced and used in the present invention. Following the lead of computational chemistry in applying multivariate data analysis techniques to the structure/property-activity relationships [see for example, Wold, et al. Multivariate data analysis in chemistry. Chemometrics-Mathematics and Statistics in Chemistry (Ed.: B. Kowalski); D. Reidel Publishing Company, Dordrecht, Holland, 1984 (ISBN 90-277-1846-6)] quantitative activity-property relationships of proteins can be derived using well-known mathematical techniques, such as statistical regression, pattern recognition and classification [see for example Norman et al. Applied Regression Analysis. Wiley-Interscience; 3rd edition (April 1998) ISBN: 0471170828; Kandel, Abraham et al. Computer-Assisted Reasoning in Cluster Analysis. Prentice Hall

PTR, (May 11, 1995), ISBN: 0133418847; Krzanowski, Wojtek. Principles of Multivariate Analysis: A User's Perspective (Oxford Statistical Science Series, No 22 (Paper)). Oxford University Press; (December 2000), ISBN: 0198507089; Witten, Ian H. et al Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann; (October 11, 1999), ISBN:1558605525; Denison David G. T. (Editor) et al Bayesian Methods for Nonlinear Classification and Regression (Wiley Series in Probability and Statistics). John Wiley & Sons; (July 2002), ISBN: 0471490369; Ghose, Arup K. et al. Combinatorial Library Design and Evaluation Principles, Software, Tools, and Applications in Drug Discovery. ISBN: 0-8247-0487-8]. The properties of proteins can be derived from empirical and theoretical models (for example, analysis of likely contact residues or calculated physicochemical property) of proteins sequence, functional and three-dimensional structures and these properties can be considered individually and in combination.

Amino acids are referred to herein using the name of the amino acid, the three-letter abbreviation or the single letter abbreviation. The term "protein", as used herein, includes proteins, polypeptides, and peptides. As used herein, the term "amino acid sequence" is synonymous with the term "polypeptide" and/or the term "protein". In some instances, the term "amino acid sequence" is synonymous with the term "peptide". The terms "protein" and "polypeptide" are used interchangeably herein. In the present disclosure and claims, the conventional one-letter and three-letter codes for amino acid residues may be used. The 3-letter code for amino acids as defined in conformity with the IUPACIUB Joint Commission on Biochemical Nomenclature (JCBN). It is also understood that a polypeptide may be coded for by more than one nucleotide sequence due to the degeneracy of the genetic code.

Amino acid residues at non-conserved positions may be substituted with conservative or non-conservative residues. In particular, conservative amino acid replacements are contemplated.

A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art, including basic side chains (e.g., lysine, arginine, or histidine), acidic side chains (e.g., aspartic acid or glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, or cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, or tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, or histidine). Thus, if an amino acid in a polypeptide is replaced with another amino acid from the same side chain family, the amino acid substitution is considered to be conservative. The inclusion of conservatively modified variants in a protein of the invention does not exclude other forms of variant, for example polymorphic variants, interspecies homologs, and alleles.

“Non-conservative amino acid substitutions” include those in which (i) a residue having an electropositive side chain (e.g., Arg, His or Lys) is substituted for, or by, an electronegative residue (e.g., Glu or Asp), (ii) a hydrophilic residue (e.g., Ser or Thr) is substituted for, or by, a hydrophobic residue (e.g., Ala, Leu, Ile, Phe or Val), (iii) a cysteine or proline is substituted for, or by, any other
5 residue, or (iv) a residue having a bulky hydrophobic or aromatic side chain (e.g., Val, His, Ile or Trp) is substituted for, or by, one having a smaller side chain (e.g., Ala or Ser) or no side chain (e.g., Gly).

“Insertions” or “deletions” are typically in the range of about 1, 2, or 3 amino acids. The variation allowed may be experimentally determined by systematically introducing insertions or deletions of amino acids in a protein using recombinant DNA techniques and assaying the resulting
10 recombinant variants for activity. This does not require more than routine experiments for a skilled person.

A “fragment” of a polypeptide comprises at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97% or more of the original polypeptide.

The polynucleotides of the present invention may be prepared by any means known in the
15 art. For example, large amounts of the polynucleotides may be produced by replication in a suitable host cell. The natural or synthetic DNA fragments coding for a desired fragment will be incorporated into recombinant nucleic acid constructs, typically DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the DNA constructs will be suitable for autonomous replication in a unicellular host, such as yeast or bacteria, but may also be intended for
20 introduction to and integration within the genome of a cultured insect, mammalian, plant or other eukaryotic cell lines.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g. by the phosphoramidite method or the tri-ester method, and may be performed on commercial automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the
25 single stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

When applied to a nucleic acid sequence, the term “isolated” in the context of the present invention denotes that the polynucleotide sequence has been removed from its natural genetic milieu
30 and is thus free of other extraneous or unwanted coding sequences (but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators), and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment.

In view of the degeneracy of the genetic code, considerable sequence variation is possible among the polynucleotides of the present invention. Degenerate codons encompassing all possible codons for a given amino acid are set forth below:

Amino Acid	Codons	Degenerate Codon
Cys	TGC TGT	TGY
Ser	AGC AGT TCA TCC TCG TCT	WSN
Thr	ACA ACC ACG ACT	ACN
Pro	CCA CCC CCG CCT	CCN
Ala	GCA GCC GCG GCT	GCN
Gly	GGA GGC GGG GGT	GGN
Asn	AAC AAT	AAY
Asp	GAC GAT	GAY
Glu	GAA GAG	GAR
Gln	CAA CAG	CAR
His	CAC CAT	CAY
Arg	AGA AGG CGA CGC CGG CGT	MGN
Lys	AAA AAG	AAR
Met	ATG	ATG
Ile	ATA ATC ATT	ATH
Leu	CTA CTC CTG CTT TTA TTG	YTN
Val	GTA GTC GTG GTT	GTN
Phe	TTC TTT	TTY
Tyr	TAC TAT	TAY
Trp	TGG	TGG
Ter	TAA TAG TGA	TRR
Asn/ Asp		RAY
Glu/ Gln		SAR
Any		NNN

5 One of ordinary skill in the art will appreciate that flexibility exists when determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, some polynucleotides encompassed by the degenerate sequence may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequences of the present invention.

10 A “variant” nucleic acid sequence has substantial homology or substantial similarity to a reference nucleic acid sequence (or a fragment thereof). A nucleic acid sequence or fragment thereof is “substantially homologous” (or “substantially identical”) to a reference sequence if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 70%, 75%, 80%, 85, 90,
15 91, 92, 93, 94, 95, 96, 97, 98, 99 or more% of the nucleotide bases. Methods for homology determination of nucleic acid sequences are known in the art.

Alternatively, a “variant” nucleic acid sequence is substantially homologous with (or substantially identical to) a reference sequence (or a fragment thereof) if the “variant” and the

reference sequence they are capable of hybridizing under stringent (e.g. highly stringent) hybridization conditions. Nucleic acid sequence hybridization will be affected by such conditions as salt concentration (e.g. NaCl), temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the
5 hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions are preferably employed, and generally include temperatures in excess of 30°C, typically in excess of 37°C and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. The pH is typically between 7.0 and 8.3. The combination of parameters is much more important than any
10 single parameter.

Methods of determining nucleic acid percentage sequence identity are known in the art. By way of example, when assessing nucleic acid sequence identity, a sequence having a defined number of contiguous nucleotides may be aligned with a nucleic acid sequence (having the same number of contiguous nucleotides) from the corresponding portion of a nucleic acid sequence of the present
15 invention. Tools known in the art for determining nucleic acid percentage sequence identity include Nucleotide BLAST (as described below).

One of ordinary skill in the art appreciates that different species exhibit “preferential codon usage”. As used herein, the term “preferential codon usage” refers to codons that are most frequently used in cells of a certain species, thus favouring one or a few representatives of the possible codons
20 encoding each amino acid. For example, the amino acid threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian host cells ACC is the most commonly used codon; in other species, different codons may be preferential. Preferential codons for a particular host cell species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into recombinant DNA can, for example, enhance
25 production of the protein by making protein translation more efficient within a particular cell type or species. Thus, according to the invention, in addition to the gag-pol genes any nucleic acid sequence may be codon-optimised for expression in a host or target cell. In particular, the vector genome (or corresponding plasmid), the REV gene (or corresponding plasmid), the fusion protein (F) gene (or correspond plasmid) and/or the hemagglutinin-neuraminidase (HN) gene (or corresponding plasmid,
30 or any combination thereof may be codon-optimised.

A “fragment” of a polynucleotide of interest comprises a series of consecutive nucleotides from the sequence of said full-length polynucleotide. By way of example, a “fragment” of a polynucleotide of interest may comprise (or consist of) at least 30 consecutive nucleotides from the sequence of said polynucleotide (e.g. at least 35, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500,

550, 600, 650, 700, 750, 800 850, 900, 950 or 1000 consecutive nucleic acid residues of said polynucleotide). A fragment may include at least one antigenic determinant and/or may encode at least one antigenic epitope of the corresponding polypeptide of interest. Typically, a fragment as defined herein retains the same function as the full-length polynucleotide.

5 The terms "decrease", "reduced", "reduction", or "inhibit" are all used herein to mean a decrease by a statistically significant amount. The terms "reduce," "reduction" or "decrease" or "inhibit" typically means a decrease by at least 10% as compared to a reference level (e.g. the absence of a given treatment) and can include, for example, a decrease by at least about 10%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99% , or more. As used herein, "reduction" or "inhibition" encompasses a complete inhibition or reduction as compared to a reference level. "Complete inhibition" is a 100% inhibition (i.e. abrogation) as compared to a reference level.

15 The terms "increased", "increase", "enhance", or "activate" are all used herein to mean an increase by a statically significant amount. The terms "increased", "increase", "enhance", or "activate" can mean an increase of at least 25%, at least 50% as compared to a reference level, for example an increase of at least about 50%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 100%, or at least about 150%, or at least about 200%, or at least about 250% or more compared with a reference level, or at least about a 1.5-fold, or at least about a 2-fold, or at least about a 2.5-fold, or at least about a 3-fold, or at least about a 4-fold, or at least about a 5-fold or at least about a 10-fold increase, or any increase between 1.5-fold and 10-fold or greater as compared to a reference level. In the context of a yield or titre, an "increase" is an observable or statistically significant increase in such level.

25 The terms "individual", "subject", and "patient", are used interchangeably herein to refer to a mammalian subject for whom diagnosis, prognosis, disease monitoring, treatment, therapy, and/or therapy optimisation is desired. The mammal can be (without limitation) a human, non-human primate, mouse, rat, dog, cat, horse, or cow. In a preferred embodiment, the individual, subject, or patient is a human. An "individual" may be an adult, juvenile or infant. An "individual" may be male or female.

30 A "subject in need" of treatment for a particular condition can be an individual having that condition, diagnosed as having that condition, or at risk of developing that condition.

 A subject can be one who has been previously diagnosed with or identified as suffering from or having a condition in need of treatment or one or more complications or symptoms related to such

a condition, and optionally, have already undergone treatment for a condition as defined herein or the one or more complications or symptoms related to said condition. Alternatively, a subject can also be one who has not been previously diagnosed as having a condition as defined herein or one or more or symptoms or complications related to said condition. For example, a subject can be one who
 5 exhibits one or more risk factors for a condition, or one or more or symptoms or complications related to said condition or a subject who does not exhibit risk factors.

As used herein, the term “healthy individual” refers to an individual or group of individuals who are in a healthy state, e.g. individuals who have not shown any symptoms of the disease, have not been diagnosed with the disease and/or are not likely to develop the disease e.g. cystic fibrosis
 10 (CF) or any other disease described herein). Preferably said healthy individual(s) is not on medication affecting CF and has not been diagnosed with any other disease. The one or more healthy individuals may have a similar sex, age, and/or body mass index (BMI) as compared with the test individual. Application of standard statistical methods used in medicine permits determination of normal levels of expression in healthy individuals, and significant deviations from such normal levels.

15 Herein the terms “control” and “reference population” are used interchangeably.

The term “pharmaceutically acceptable” as used herein means approved by a regulatory agency of the Federal or a state government, or listed in the U.S. Pharmacopeia, European Pharmacopeia or other generally recognized pharmacopeia

The publications discussed herein are provided solely for their disclosure prior to the filing
 20 date of the present application. Nothing herein is to be construed as an admission that such publications constitute prior art to the claims appended hereto.

Disclosure related to the various methods of the invention are intended to be applied equally to other methods, therapeutic uses or methods, the data storage medium or device, the computer program product, and *vice versa*.

25

Retroviral and Lentiviral vectors

The invention relates to the production of a retroviral/lentiviral (e.g. SIV) construct. The term “retrovirus” refers to any member of the Retroviridae family of RNA viruses that encode the enzyme reverse transcriptase. The term “lentivirus” refers to a family of retroviruses. Examples of retroviruses
 30 suitable for use in the present invention include gammaretroviruses such as murine leukaemia virus (MLV) and feline leukaemia virus (FLV). Examples of lentiviruses suitable for use in the present invention include Simian immunodeficiency virus (SIV), Human immunodeficiency virus (HIV), Feline immunodeficiency virus (FIV), Equine infectious anaemia virus (EIAV), and Visna/maedi virus. Preferably the invention relates to lentiviral vectors and the production thereof. A particularly

preferred lentiviral vector is an SIV vector (including all strains and subtypes), such as a SIV-AGM (originally isolated from African green monkeys, *Cercopithecus aethiops*). Alternatively the invention relates to HIV vectors.

The retroviral/lentiviral (e.g. SIV) vectors of the present invention are typically pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus. Preferably the respiratory paramyxovirus is a Sendai virus (murine parainfluenza virus type 1). The retroviral/lentiviral (e.g. SIV) vectors of the present invention may be pseudotyped with proteins from another virus, provided that the use of codon-optimised gag-pol genes (e.g. from SIV) does not negatively impact the manufactured titre of the vector, or even results in an increased titre of the vector. Non-limiting examples of other proteins that may be used to pseudotype retroviral/lentiviral (e.g. SIV) vectors of the present invention include G glycoprotein from Vesicular Stomatitis Virus (G-VSV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein or modified forms thereof; such as those described in UK Patent Application Nos. 2118685.3 and 2105278.2, each of which is herein incorporated by reference in its entirety. Thus, the invention may relate to the production of SIV pseudotyped with G-VSV or SIV pseudotyped with a SARS-CoV-2 spike protein, using codon-optimised gag-pol genes.

A retroviral/lentiviral (e.g. SIV) vector produced according to the invention may be integrase-competent (IC). Alternatively, the lentiviral (e.g. SIV) vector may be integrase-deficient (ID).

Retroviral/Lentiviral vectors, such as those produced according to the invention, can integrate into the genome of transduced cells and lead to long-lasting expression, making them suitable for transduction of stem/progenitor cells. In the lung, several cell types with regenerative capacity have been identified as responsible for maintaining specific cell lineages in the conducting airways and alveoli. These include basal cells and submucosal gland duct cells in the upper airways, club cells and neuroendocrine cells in the bronchiolar airways, bronchioalveolar stem cells in the terminal bronchioles and type II pneumocytes in the alveoli. Therefore, and without being bound by theory, it is believed that said retroviral/lentiviral (e.g. SIV) vectors bring about long term gene expression of the transgene of interest by introducing the transgene into one or more long-lived airway epithelial cells or cell types, such as basal cells and submucosal gland duct cells in the upper airways, club cells and neuroendocrine cells in the bronchiolar airways, bronchioalveolar stem cells in the terminal bronchioles and type II pneumocytes in the alveoli.

Accordingly, the retroviral/lentiviral (e.g. SIV) vectors produced according to the invention may transduce one or more cells or cell lines with regenerative potential within the lung (including the airways and respiratory tract) to achieve long term gene expression. For example, the retroviral/lentiviral (e.g. SIV) vectors may transduce basal cells, such as those in the upper

airways/respiratory tract. Basal cells have a central role in processes of epithelial maintenance and repair following injury. In addition, basal cells are widely distributed along the human respiratory epithelium, with a relative distribution ranging from 30% (larger airways) to 6% (smaller airways).

The retroviral/lentiviral (e.g. SIV) vectors produced according to the invention may be used to transduce isolated and expanded stem/progenitor cells *ex vivo* prior administration to a patient. Preferably, the retroviral/lentiviral (e.g. SIV) vectors produced according to the invention are used to transduce cells within the lung (or airways/respiratory tract) *in vivo*.

The retroviral/lentiviral (e.g. SIV) vectors of the invention demonstrate remarkable resistance to shear forces with only modest reduction in transduction ability when passaged through clinically-relevant delivery devices such as bronchoscopes, spray bottles and nebulisers.

The retroviral/lentiviral (e.g. SIV) vectors of the present invention enable high levels of transgene expression, resulting in high levels (therapeutic levels) of expression of a therapeutic protein. The retroviral/lentiviral (e.g. SIV) vectors of the present invention typically provide high expression levels of a transgene when administered to a patient. The terms high expression and therapeutic expression are used interchangeably herein. Expression may be measured by any appropriate method (qualitative or quantitative, preferably quantitative), and concentrations given in any appropriate unit of measurement, for example ng/ml or nM.

Expression of a transgene of interest may be given relative to the expression of the corresponding endogenous (defective) gene in a patient. Expression may be measured in terms of mRNA or protein expression. The expression of the transgene of the invention, such as a functional CFTR gene, may be quantified relative to the endogenous gene, such as the endogenous (dysfunctional) CFTR genes in terms of mRNA copies per cell or any other appropriate unit.

Expression levels of a transgene and/or the encoded therapeutic protein of the invention may be measured in the lung tissue, epithelial lining fluid and/or serum/plasma as appropriate. A high and/or therapeutic expression level may therefore refer to the concentration in the lung, epithelial lining fluid and/or serum/plasma.

The transgene included in the vector of the invention may be modified to facilitate expression. For example, the transgene sequence may be in CpG-depleted (or CpG-free) and/or codon-optimised form to facilitate gene expression. Standard techniques for modifying the transgene sequence in this way are known in the art.

The retroviral/lentiviral (e.g. SIV) vectors of the invention exhibit efficient airway cell uptake, enhanced transgene expression, and suffer no loss of efficacy upon repeated administration. Accordingly, the retroviral/lentiviral (e.g. SIV) vectors of the invention are capable of producing long-lasting, repeatable, high-level expression in airway cells without inducing an undue immune response.

The retroviral/lentiviral (e.g. SIV) vectors of the present invention enable long-term transgene expression, resulting in long-term expression of a therapeutic protein. As described herein, the phrases “long-term expression”, “sustained expression”, “long-lasting expression” and “persistent expression” are used interchangeably. Long-term expression according to the present invention
5 means expression of a therapeutic gene and/or protein, preferably at therapeutic levels, for at least 45 days, at least 60 days, at least 90 days, at least 120 days, at least 180 days, at least 250 days, at least 360 days, at least 450 days, at least 730 days or more. Preferably long-term expression means expression for at least 90 days, at least 120 days, at least 180 days, at least 250 days, at least 360 days, at least 450 days, at least 720 days or more, more preferably at least 360 days, at least 450 days, at
10 least 720 days or more. This long-term expression may be achieved by repeated doses or by a single dose.

Repeated doses may be administered twice-daily, daily, twice-weekly, weekly, monthly, every two months, every three months, every four months, every six months, yearly, every two years, or more. Dosing may be continued for as long as required, for example, for at least six months, at least
15 one year, two years, three years, four years, five years, ten years, fifteen years, twenty years, or more, up to for the lifetime of the patient to be treated.

The retroviral/lentiviral (e.g. SIV) vector comprises a promoter operably linked to a transgene, enabling expression of the transgene. Typically the promoter is a hybrid human CMV enhancer/EF1a (hCEF) promoter. This hCEF promoter may lack the intron corresponding to nucleotides 570-709 and
20 the exon corresponding to nucleotides 728-733 of the hCEF promoter. A preferred example of an hCEF promoter sequence of the invention is provided by SEQ ID NO: 10. The promoter may be a CMV promoter. An example of a CMV promoter sequence is provided by SEQ ID NO: 11. The promoter may be a human elongation factor 1a (EF1a) promoter. An example of a EF1a promoter is provided by SEQ ID NO: 12. Other promoters for transgene expression are known in the art and their suitability for the
25 retroviral/lentiviral (e.g. SIV) vectors of the invention determined using routine techniques known in the art. Non-limiting examples of other promoters include UbC and UCOE. As described herein, the promoter may be modified to further regulate expression of the transgene of the invention.

The promoter included in the retroviral/lentiviral (e.g. SIV) vector of the invention may be specifically selected and/or modified to further refine regulation of expression of the therapeutic
30 gene. Again, suitable promoters and standard techniques for their modification are known in the art. As a non-limiting example, a number of suitable (CpG-free) promoters suitable for use in the present invention are described in Pringle *et al.* (J. Mol. Med. Berl. 2012, 90(12): 1487-96), which is herein incorporated by reference in its entirety. Preferably, the retroviral/lentiviral vectors (particularly SIV F/HN vectors) of the invention comprise a hCEF promoter having low or no CpG dinucleotide content.

The hCEF promoter may have all CG dinucleotides replaced with any one of AG, TG or GT. Thus, the hCEF promoter may be CpG-free. A preferred example of a CpG-free hCEF promoter sequence of the invention is provided by SEQ ID NO: 10. The absence of CpG dinucleotides further improves the performance of retroviral/lentiviral (e.g. SIV) vectors of the invention and in particular in situations where it is not desired to induce an immune response against an expressed antigen or an inflammatory response against the delivered expression construct. The elimination of CpG dinucleotides reduces the occurrence of flu-like symptoms and inflammation which may result from administration of constructs, particularly when administered to the airways.

The retroviral/lentiviral (e.g. SIV) vector of the invention may be modified to allow shut down of gene expression. Standard techniques for modifying the vector in this way are known in the art. As a non-limiting example, Tet-responsive promoters are widely used.

Preferably, the invention relates to F/HN retroviral/lentiviral vectors comprising a promoter and a transgene, particularly SIV F/HN vectors. The F/HN pseudotyping is particularly efficient at targeting cells in the airway epithelium, and as such, for therapeutic applications it is typically delivered to cells of the respiratory tract, including the cells of the airway epithelium. Accordingly, the retroviral/lentiviral (e.g. SIV) vectors of the invention are particularly suited for treatment of diseases or disorders of the airways, respiratory tract, or lung. Typically, the retroviral/lentiviral (e.g. SIV) vectors may be used for the treatment of a genetic respiratory disease.

A retroviral/lentiviral (e.g. SIV) vector of the invention may comprise a transgene that encodes a polypeptide or protein that is therapeutic for the treatment of such diseases, particularly a disease or disorder of the airways, respiratory tract, or lung.

Accordingly, a retroviral/lentiviral (e.g. SIV) vector of the invention may comprise a transgene encoding a protein selected from: (i) a secreted therapeutic protein, optionally Alpha-1 Antitrypsin (A1AT), Factor VIII, Surfactant Protein B (SFTPB), Factor VII, Factor IX, Factor X, Factor XI, von Willebrand Factor, Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) and a monoclonal antibody against an infectious agent; or (ii) *CFTR*, *ABCA3*, *DNAH5*, *DNAH11*, *DNAI1*, and *DNAI2*. Other examples of transgenes that may be comprised in a retroviral/lentiviral (e.g. SIV) vector of the invention include genes related to or associated with other surfactant deficiencies.

Preferably, the transgene encodes a CFTR. An example of a *CFTR* cDNA is provided by SEQ ID NO: 13. Variants thereof (as described therein) are also included, particularly variants with at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100% to SEQ ID NO: 13).

The transgene may encode an A1AT. An example of an *A1AT* transgene is provided by SEQ ID NO: 14, or by the complementary sequence of SEQ ID NO: 15. SEQ ID NO: 14 is a codon-optimized CpG depleted *A1AT* transgene previously designed by the present inventors to enhance translation in

human cells. Such optimisation has been shown to enhance gene expression by up to 15-fold. Variants of same sequence (as defined herein) which possess the same technical effect of enhancing translation compared with the unmodified (wild-type) *A1AT* gene sequence are also encompassed by the present invention. The polypeptide encoded by said *A1AT* transgene, may be exemplified by the polypeptide of SEQ ID NO: 16. Variants thereof (as described therein) are also included, particularly variants with at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100% to SEQ ID NO: 14, 15 or 16.

The transgene may encode a *FVIII*. Examples of a *FVIII* transgene are provided by SEQ ID NOs: 17 and 18, or by the respective complementary sequences of SEQ ID NO: 19 and 20. The polypeptide encoded by the *FVIII* transgene, may be exemplified by the polypeptide of SEQ ID NO: 21 or 22. Variants thereof (as described therein) are also included, particularly variants with at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100% to any one of SEQ ID NOs: 17 to 22.

The transgene of the invention may be any one or more of *DNAH5*, *DNAH11*, *DNAI1*, and *DNAI2*, or other known related gene.

When the respiratory tract epithelium is targeted for delivery of the retroviral/lentiviral (e.g. SIV) vector, the transgene may encode *A1AT*, *SFTPB*, or *GM-CSF*. The transgene may encode a monoclonal antibody (mAb) against an infectious agent. The transgene may encode anti-TNF alpha. The transgene may encode a therapeutic protein implicated in an inflammatory, immune or metabolic condition.

A retroviral/lentiviral (e.g. SIV) vector of the invention may be delivered to the cells of the respiratory tract to allow production of proteins to be secreted into circulatory system. In such embodiments, the transgene may encode for Factor VII, Factor VIII, Factor IX, Factor X, Factor XI and/or von Willebrand's factor. Such a vector may be used in the treatment of diseases, particularly cardiovascular diseases and blood disorders, preferably blood clotting deficiencies such as haemophilia. Again, the transgene may encode an mAb against an infectious agent or a protein implicated in an inflammatory, immune or metabolic condition, such as, lysosomal storage disease.

The retroviral/lentiviral (e.g. SIV) vector of the invention may have no intron positioned between the promoter and the transgene. Similarly, there may be no intron between the promoter and the transgene in the vector genome (pDNA1) plasmid (for example, pGM326 as described herein, illustrated in Figure 2A and with the sequence of SEQ ID NO: 3).

In some preferred embodiments, the retroviral/lentiviral (e.g. SIV) vector comprises a hCEF promoter and a *CFTR* transgene, including those described herein. Optionally said retroviral/lentiviral (e.g. SIV) vector may have no intron positioned between the promoter and the transgene. Such a retroviral/lentiviral (e.g. SIV) vector may be produced by the method described herein, using a genome plasmid carrying the *CFTR* transgene and a promoter.

In some preferred embodiments, the retroviral/lentiviral (e.g. SIV) vector comprises a hCEF promoter and an A1AT transgene, including those described herein. Optionally said retroviral/lentiviral (e.g. SIV) vector may have no intron positioned between the promoter and the transgene. Such a retroviral/lentiviral (e.g. SIV) vector may be produced by the method described
5 herein, using a genome plasmid carrying the A1AT transgene and a promoter.

In some preferred embodiments, the retroviral/lentiviral (e.g. SIV) vector comprises a hCEF or CMW promoter and an FVIII transgene, including those described herein. Optionally said retroviral/lentiviral (e.g. SIV) vector may have no intron positioned between the promoter and the transgene. Such a retroviral/lentiviral (e.g. SIV) vector may be produced by the method described
10 herein, using a genome plasmid carrying the FVIII transgene and a promoter.

The retroviral/lentiviral (e.g. SIV) vector as described herein comprises a transgene. The transgene comprises a nucleic acid sequence encoding a gene product, e.g., a protein, particularly a therapeutic protein.

For example, in one embodiment, the nucleic acid sequence encoding a CFTR, A1AT or FVIII
15 comprises (or consists of) a nucleic acid sequence having at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100%) sequence identity to the CFTR, A1AT or FVIII nucleic acid sequence respectively, examples of which are described herein. In a further embodiment, the nucleic acid sequence encoding CFTR, A1AT or FVIII comprises (or consists of) a nucleic acid sequence having at least 95% (such as at least 95, 96, 97, 98, 99 or 100%) sequence identity to the CFTR, A1AT or FVIII
20 nucleic acid sequence respectively, examples of which are described herein. In one embodiment, the nucleic acid sequence encoding CFTR is provided by SEQ ID NO: 13, the nucleic acid sequence encoding A1AT is provided by SEQ ID NO: 14, or by the complementary sequence of SEQ ID NO: 15 and/or the nucleic acid sequence encoding FVIII is provided by SEQ ID NO: 17 and 18, or by the respective complementary sequences of SEQ ID NO: 19 and 20, or variants thereof.

25 The amino acid sequence of the CFTR, A1AT or FVIII transgene may comprise (or consist of) an amino acid sequence having at least 95% (such as at least 95, 96, 97, 98, 99 or 100%) sequence identity to the functional CFTR, A1AT or FVIII polypeptide sequence respectively.

The retroviral/lentiviral (e.g. SIV) vectors of the invention may comprise a central polypurine tract (cPPT) and/or the Woodchuck hepatitis virus posttranscriptional regulatory elements (WPRE). An
30 exemplary WPRE sequence is provided by SEQ ID NO: 23.

Methods of Production

As described herein, the present inventors have demonstrated for the first time that the use of codon-optimised gal-pol genes from SIV does not negatively impact the manufactured titre of a SIV

vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, and can even result in an increased titre of the vector. In addition, the inventors have further shown that the use of codon-optimised gag-pol genes can be further combined with the use of a modified vector genome plasmid as described herein whilst maintaining, or even
5 increasing the vector titre.

Codon optimisation is a technique to maximise protein expression by increasing the translational efficiency of the encoding gene. Translational efficiency is increased by modification of the nucleic acid sequence. Codon optimisation is routine in the art, and it is within the routine practice of one of ordinary skill to devise a codon-optimised version of a given nucleic acid sequence. However,
10 what is not straightforward is predicting the effect of codon optimisation on other parameters. For example, as described herein, conventional wisdom teaches that under normal manufacturing conditions (when the vector genome plasmid, rather than the gag-pol genes, is limiting), codon-optimisation of the gag-pol genes typically decreases vector yield.

Accordingly, the present invention provides a method of producing a retroviral/lentiviral (e.g.
15 SIV) vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, and which comprises a promoter and a transgene, wherein said method comprises the use of codon-optimised gag-pol genes. Preferably said vector is a lentiviral vector, with Simian immunodeficiency virus (SIV) vectors being particularly preferred.

Typically the codon-optimised gag-pol genes used in the production methods of the invention
20 are matched to the retroviral/lentiviral vector being produced. By way of non-limiting example, when the lentiviral vector is an HIV vector, the codon-optimised gag-pol genes used in the production methods of the invention are HIV gag-pol genes. By way of non-limiting example, when the lentiviral vector is an SIV vector, the codon-optimised gag-pol genes used in the production methods of the invention are SIV gag-pol genes.

Preferably the codon-optimised gag-pol genes used in the production methods of the invention are SIV gag-pol genes. Exemplary wild-type SIV gag-pol genes that may be modified to produce codon-optimised gag-pol genes are given in SEQ ID NO: 2. The modifications made to the wild-type gag-pol genes of SEQ ID NO: 2 in order to arrive at an exemplary codon-optimised gag-pol genes of the invention (SEQ ID NO: 1) are shown in the alignment in Figure 1.
25

In addition to codon-optimisation, the codon-optimised gag-pol genes used in the production methods of the invention may comprise other modifications, such as a translational slip (which allows translation to slip from one region to another to allow the production of both Gag and Pol). Any suitable variation of codon usage may be used in the codon-optimised gag-pol genes of the invention, provided that (i) homology between the vector genome plasmid and GagPol plasmid is reduced to
30

minimise the risk of RCL production and (ii) after codon optimisation there is production of sufficient GagPol without the inclusion of RRE (this further reduces homology and the risk of RCL production).

The codon-optimised gag-pol genes used in the production methods of the invention may be completely (100%) or partially codon-optimised. Partial codon-optimisation encompasses at least
5 70%, at least 80%, at least 95%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more codon optimisation.

Preferably, the gag-pol genes themselves are completely codon-optimised, but may comprise non-contain regions of non-codon-optimised sequence (e.g. between the gag and pol genes). By way of non-limiting example, to maintain the translational slip of reading frames between the gag and pol
10 genes, the region around the translational slip sequence may not be codon-optimised (e.g. in case the precise translational slip sequence is important for this function). A non-codon-optimised translational slip sequence within codon-optimised gag-pol genes is exemplified in SEQ ID NO: 1.

Preferably, the codon-optimised gag-pol genes used in a method of the invention comprise or consist of the nucleic acid sequence of SEQ ID NO: 1, or a variant thereof (as defined herein). In
15 particular, the codon-optimised gag-pol genes used in a method of the invention comprise or consist of a nucleic acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to SEQ ID NO: 1. Preferably, the codon-optimised gag-pol genes used in a method of the invention comprise or consist of a nucleic acid sequence having at least 90%, more preferably at least 95%, even more preferably at least 98%, or
20 more sequence identity to SEQ ID NO: 1. The codon-optimised gag-pol genes of SEQ ID NO: 1 comprise a translational slip, and so do not form a single conventional open reading frame.

The method of the invention may be a scalable GMP-compatible method. Thus, the method of the invention typically allows the generation of high titre purified F/HN retroviral/lentiviral (e.g. SIV) vectors. Typically a method of the invention produces a titre of retroviral/lentiviral (e.g. SIV) vector
25 that is at least equivalent to the titre of retroviral/lentiviral (e.g. SIV) vector produced by a corresponding method which does not use codon-optimised gag-pol genes. As used herein, the term "equivalent" may be defined such that the use of the codon-optimised gag-pol genes does not significantly decrease the titre of retroviral/lentiviral (e.g. SIV) vector compared with the use of the corresponding non-codon-optimised gag-pol genes. By way of non-limiting example, a method of the
30 invention produces a titre of retroviral/lentiviral (e.g. SIV) vector that is no more than 2-fold lower, no more than 1.5-fold lower, no more than 1.0-fold lower, no more than 0.5-fold lower, no more than 0.25-fold lower, or less than the titre of retroviral/lentiviral (e.g. SIV) vector compared with the use of the corresponding non-codon-optimised gag-pol genes. The term "equivalent" may be defined such that titre of retroviral/lentiviral (e.g. SIV) vector produced by a method using codon-optimised gag-

pol genes is statistically unchanged (e.g. $p < 0.05$, $p < 0.01$) compared with the titre of retroviral/lentiviral (e.g. SIV) vector produced by a method using the corresponding non-codon-optimised gag-pol genes.

Preferably, a method of the invention produces a titre of retroviral/lentiviral (e.g. SIV) vector
5 that is increased compared with the titre of retroviral/lentiviral (e.g. SIV) vector produced by a corresponding method which does not use codon-optimised gag-pol genes. The titre of retroviral/lentiviral (e.g. SIV) vector may be at least 1.5-fold, at least 2-fold, or at least 2.5-fold greater than the titre of retroviral/lentiviral (e.g. SIV) vector produced by a corresponding method which does not use codon-optimised gag-pol genes.

10 The production of retroviral/lentiviral (e.g. SIV) vectors typically employs one or more plasmids which provide the elements needed for the production of the vector: the genome for the retroviral/lentiviral vector, the Gag-Pol, Rev, F and HN. Multiple elements can be provided on a single plasmid. Preferably each element is provided on a separate plasmid, such that there five plasmids, one for each of the vector genome, the Gag-Pol, Rev, F and HN, respectively.

15 Alternatively, a single plasmid may provide the Gag-Pol and Rev elements, and may be referred to as a packaging plasmid (pDNA2). The remaining elements (genome, F and HN) may be provided by separate plasmids (pDNA1, pDNA3a, pDNA3b respectively), such that four plasmids are used for the production of a retroviral/lentiviral (e.g. SIV) vector according to the invention. In the four plasmid methods, pDNA1, pDNA3a and pDNA3b may be as described herein in the context of the
20 five-plasmid method.

Preferably, the codon-optimised gag-pol genes used in a method of the invention are comprised in a plasmid that comprises or consists of a nucleic acid sequence of SEQ ID NO: 5 (pGM691), or a variant thereof (as defined herein). In particular, the codon-optimised gag-pol genes used in a method of the invention are comprised in a plasmid that comprises or consists of a nucleic acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%,
25 at least 98%, at least 99% or more sequence identity to SEQ ID NO: 5. Preferably, the codon-optimised gag-pol genes used in a method of the invention are comprised in a plasmid that comprises or consists of a nucleic acid sequence having at least 90%, more preferably at least 95%, even more preferably at least 98%, or more sequence identity to SEQ ID NO: 5. In the plasmid of SEQ ID NO: 5 (or variants
30 thereof): (i) the codon-optimised gag-pol genes of SEQ ID NO: 1 comprise a translational slip, and so do not form a single conventional open reading frame; and (ii) the codon-optimised gag-pol genes of SEQ ID NO: 1 are operably linked to a CAG promoter.

In the preferred five plasmid method of the invention, the vector genome plasmid encodes all the genetic material that is packaged into final retroviral/lentiviral vector, including the transgene.

Typically only a portion of the genetic material found in the vector genome plasmid ends up in the virus. The vector genome plasmid may be designated herein as “pDNA1”, and typically comprises the transgene and the transgene promoter.

5 The other four plasmids are manufacturing plasmids encoding the Gag-Pol, Rev, F and HN proteins. These plasmids may be designated “pDNA2a”, “pDNA2b”, “pDNA3a” and “pDNA3b” respectively.

Modifications may be made to the vector genome plasmid (pDNA1), particularly to further improve the safety profile of the vector. As exemplified herein, such modifications may comprise or consist of modifying the pDNA1 sequence to remove viral, particularly retroviral/lentiviral (e.g. SIV),
10 ORFs from the pDNA1 sequence. Thus, the methods of the invention may use a modified pDNA1 which comprises a reduced number of non-transgene ORFs. Said modified pDNA1 may comprise modifications within any region of the plasmid sequence. In particular, a modified pDNA1 may comprise modifications to remove: (i) 5' to 3' ORFs; (ii) ORFs of ≥ 100 amino acids; and/or (iii) ORFs upstream of the transgene and/or the promoter operably linked to the transgene. Whilst a modified
15 pDNA1 may comprise no ORFs other than the transgene, this is not essential. Rather, a modified pDNA1 may still comprise ORFs other than the transgene, but may comprise a reduced number of non-transgene ORFs compared to the unmodified pDNA1 from which it is derived. By way of non-limiting example, a modified pDNA1 may comprise at least 1, at least 2, at least 3, at least 4, at least 5 or more fewer non-transgene ORFs compared with the corresponding unmodified pDNA1. As a
20 specific example, pGM830 (which is derived from pGM326) comprises 2 fewer non-transgene ORFs compared with pGM326. A modified pDNA1 may comprise at least 1, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, or more modifications (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 modifications) compared with the corresponding unmodified pDNA1. By way of non-limiting example, a modified pDNA1 may comprise between about 1 to about 20, such as between
25 about 5 to about 15, or between about 5 to about 10 modifications compared with the corresponding unmodified pDNA1. As a specific example, pGM830 (which is derived from pGM326) comprises 7 modifications compared with pGM326.

As exemplified herein, the use of the pGM380 as plasmid pDNA1 has the potential to produce an improved SIV titre compared with a production method in which the pDNA1 plasmid is pGM326
30 (Figure 11), but in which all other plasmids and method parameters are kept constant. In other words, use of a modified pDNA1 such as pGM830 does not negatively impact the improved titre achieved using codon-optimised gal-pol genes, and can even potentially provide a further improvement in titre over and above the effect of using codon-optimised gal-pol genes, such as those provided by using

pGM691 as pDNA2a. The term "increased titre" as defined herein applies equally to methods of the invention which use both codon-optimised gal-pol genes and a modified pDNA1.

Typically, the lentivirus is SIV, such as SIV1, preferably SIV-AGM. The F and HN proteins are derived from a respiratory paramyxovirus, preferably a Sendai virus.

5 In a specific embodiment relating to *CFTR*, the five plasmids are characterised by Figures 2A-2F, thus pDNA1 is the pGM326 plasmid of Figure 2A or the pGM830 plasmid of Figure 2B, pDNA2a is the pGM691 plasmid of Figure 2C, pDNA2b is the pGM299 plasmid of Figure 2D, pDNA3a is the pGM301 plasmid of Figure 2E and pDNA3b is the pGM303 plasmid of Figure 2F, or variants thereof any of these plasmids (as described herein). In this embodiment, the final CFTR containing
10 retroviral/lentiviral vector may be referred to as vGM195 (see the Examples). The pGM691 plasmid and the vGM195 vector are preferred embodiments of the invention.

As exemplified herein, the use of the pGM691 as plasmid pDNA2a has the potential to produce an improved SIV titre compared with a production method in which the pDNA2a plasmid is pGM297 (Figure 2G), but in which all other plasmids and method parameters are kept constant.

15 When a method of the invention is used to produce A1AT, the five plasmids may be characterised by Figure 3 (thus plasmid pDNA1 may be pGM407) and all of Figures 2C-F (as above for the specific CFTR embodiment), or variants of any of these plasmids (as described herein).

When a method of the invention is used to produce FVIII, the five plasmids may be characterised by one of Figures 4AD (thus plasmid pDNA1 may be pGM411, pGM412, pGM413 or
20 pGM414) and all of Figures 2C-F, or variants of any of these plasmids (as described herein).

The plasmid as defined in Figure 2A is represented by SEQ ID NO: 3; the plasmid as defined in Figure 2B is represented by SEQ ID NO: 4; the plasmid as defined in Figure 2C is represented by SEQ ID NO: 5; the plasmid as defined in Figure 2D is represented by SEQ ID NO: 6; the plasmid as defined in Figure 2E is represented by SEQ ID NO: 7; the plasmid as defined in Figure 2F is represented by SEQ
25 ID NO: 8; the plasmid as defined in Figure 2G is represented by SEQ ID NO: 9; the plasmid as defined in Figure 3 is represented by SEQ ID NO: 24 and the F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-co and/or F/HN-SIV-hCEF-HFVIII-N6-co plasmids as defined in Figures 4A to 4D are represented by SEQ ID NOs: 25 to 28 respectively. Variants (as defined herein) of these plasmids are also encompassed by the present invention. In particular, variants having at least 90%
30 (such as at least 90, 92, 94, 95, 96, 97, 98, 99, 99.5 or 100%) sequence identity to any one of SEQ ID NOs: 3 to 9, 24 and 25 to 28 are encompassed.

In the five-plasmid method of the invention all five plasmids contribute to the formation of the final retroviral/lentiviral (e.g. SIV) vector. During manufacture of the retroviral/lentiviral (e.g. SIV) vector, the vector genome plasmid (pDNA1) provides the enhancer/promoter, Psi, RRE, cPPT, mWPPE,

SIN LTR, SV40 polyA (see Figure 2A or 2B), which are important for virus manufacture. Using pGM326 or pGM830 as non-limiting examples of a pDNA1, the CMV enhancer/promoter, SV40 polyA, colE1 Ori and KanR are involved in manufacture of the retroviral/lentiviral (e.g. SIV) vector of the invention (e.g. vGM195 or vGM244), but are not found in the final retroviral/lentiviral (e.g. SIV) vector. The RRE, cPPT (central polypurine tract), hCEF, soCFTR2 (transgene) and mWPRE from pGM326 or pGM830 are found in the final retroviral/lentiviral (e.g. SIV) vector. SIN LTR (long terminal repeats, SIN/IN self-inactivating) and Psi (packaging signal) may be found in the final retroviral/lentiviral (e.g. SIV) vector.

For other retroviral/lentiviral (e.g. SIV) vectors of the invention, corresponding elements from the other vector genome plasmids (pDNA1) are required for manufacture (but not found in the final vector), or are present in the final retroviral/lentiviral (e.g. SIV) vector.

The F and HN proteins from pDNA3a and pDNA3b (preferably Sendai F and HN proteins) are important for infection of target cells with the final retroviral/lentiviral (e.g. SIV) vector, i.e. for entry of a patient's epithelial cells (typically lung or nasal cells as described herein). The products of the pDNA2a and pDNA2b plasmids are important for virus transduction, i.e. for inserting the retroviral/lentiviral (e.g. SIV) DNA into the host's genome. The promoter, regulatory elements (such as WPRE) and transgene are important for transgene expression within the target cell(s).

A method of the invention may comprise or consist of the following steps: (a) growing cells in suspension; (b) transfecting the cells with one or more plasmids; (c) adding a nuclease; (d) harvesting the lentivirus (e.g. SIV); (e) adding trypsin; and (f) purification of the lentivirus (e.g. SIV).

This method may use the four- or five-plasmid system described herein. Thus, for the preferred five-plasmid method, the one or more plasmids may comprise or consist of: a vector genome plasmid pDNA1; a co-gagpol plasmid, pDNA2a; a Rev plasmid, pDNA2b; a fusion (F) protein plasmid, pDNA3a; and a hemagglutinin-neuraminidase (HN) plasmid, pDNA3b. The pDNA1 may be selected from pGM326 and pGM830, preferably pGM830. The pDNA2a may be pGM691. The pDNA2b may be pGM299. The pDNA3a may be pGM301. The pDNA3b may be pGM303. Any combination of pDNA1, pDNA2a, pDNA2b, pDNA3a and pDNA3b may be used. Preferably, the pDNA1 is pGM326 or pGM830 (pGM830 being particularly preferred); the pDNA2a is pGM691; the pDNA2b is pGM299; the pDNA3a is pGM301; and the pDNA3b is pGM303. A SIV vector produced using pGM830, pGM691, pGM299, pGM301, and pGM303 is designated vGM244. A SIV vector produced using pGM326, pGM691, pGM299, pGM301, and pGM303 is designated vGM195.

Any appropriate ratio of vector genome plasmid: co-gagpol plasmid: Rev plasmid: F plasmid: HN plasmid may be used to further optimise (increase) the retroviral/lentiviral (e.g. SIV) titre produced. By way of non-limiting example, the ratio of vector genome plasmid: co-gagpol plasmid: Rev plasmid: F plasmid: HN plasmid may be in the range of 10-40:-4-20:3-12:3-12:3-12, typically 15-

20:7-11:4-8:4-8:4-8, such as about 18-22:7-11:4-8:4-8:4-8, 19-21:8-10:5-7:5-7:5-7. Preferably the ratio of vector genome plasmid: co-gagpol plasmid: Rev plasmid: F plasmid: HN plasmid is about 20:9:6:6:6.

Steps (a)-(f) of the method are typically carried out sequentially, starting at step (a) and continuing through to step (f). The method may include one or more additional step, such as
5 additional purification steps, buffer exchange, concentration of the retroviral/lentiviral (e.g. SIV) vector after purification, and/or formulation of the retroviral/lentiviral (e.g. SIV) vector after purification (or concentration). Each of the steps may comprise one or more sub-steps. For example, harvesting may involve one or more steps or sub-steps, and/or purification may involve one or more steps or sub-steps.

10 Any appropriate cell type may be transfected with the one or more plasmids (e.g. the five-plasmids described herein) to produce a retroviral/lentiviral (e.g. SIV) vector of the invention. Typically mammalian cells, particularly human cell lines are used. Non-limiting examples of cells suitable for use in the methods of the invention are HEK293 cells (such as HEK293F or HEK293T cells) and 293T/17 cells. Commercial cell lines suitable for the production of virus are also readily available (e.g. Gibco
15 Viral Production Cells – Catalogue Number A35347 from ThermoFisher Scientific).

The cells may be grown in animal-component free media, including serum-free media. The cells may be grown in a media which contains human components. The cells may be grown in a defined media comprising or consisting of synthetically produced components.

Any appropriate transfection means may be used according to the invention. Selection of
20 appropriate transfection means is within the routine practice of one of ordinary skill in the art. By way of non-limiting example, transfection may be carried out by the use of PEIPro™, Lipofectamine2000™ or Lipofectamine3000™.

Any appropriate nuclease may be used according to the invention. Selection of appropriate nuclease is within the routine practice of one of ordinary skill in the art. Typically the nuclease is an
25 endonuclease. By way of non-limiting example, the nuclease may be Benzonase® or Denarase®. The addition of the nuclease may be at the pre-harvest stage or at the post-harvest stage, or between harvesting steps.

The trypsin activity may preferably be provided by an animal origin free, recombinant enzyme such as TrypLE Select™. The addition of trypsin may be at the pre-harvest stage or at the post-harvest
30 stage, or between harvesting steps.

Any appropriate purification means may be used to purify the retroviral/lentiviral (e.g. SIV) vector. Non-limiting examples of suitable purification steps include depth/end filtration, tangential flow filtration (TFF) and chromatography. The purification step typically comprises at least one chromatography step. Non-limiting examples of chromatography steps that may be used in

accordance with the invention include mixed-mode size exclusion chromatography (SEC) and/or anion exchange chromatography. Elution may be carried out with or without the use of a salt gradient, preferably without.

5 This method may be used to produce the retroviral/lentiviral (e.g. SIV) vectors of the invention, such as those comprising a *CFTR*, *A1AT* and/or *FVIII* gene as described herein. Alternatively, the retroviral/lentiviral (e.g. SIV) vector of the invention comprises any of the above-mentioned genes, or the genes encoding the above-mentioned proteins.

10 The method of the invention, may use any combination of one or more of the specific plasmid constructs provided by Figures 2A-2F, Figure 3 and/or Figure 4A-4D is used to provide a retroviral/lentiviral (e.g. SIV) vector of the invention. Particularly the plasmid constructs of Figures 2C-2F are used, preferably in combination with the plasmid of Figure 2B, Figure 2A, Figure 3 or Figure 4A-4D, with the plasmid of Figure 2B being particularly preferred.

15 The invention also provides codon-optimised SIV gag-pol genes. These codon-optimised SIV gag-pol genes are typically suitable for use in the methods of the invention. The codon-optimised gag-pol genes of the invention may comprise or consist of the nucleic acid sequence of SEQ ID NO: 1, or a variant thereof (as defined herein). In particular, the codon-optimised gag-pol genes of the invention may comprise or consist of a nucleic acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to SEQ ID NO: 1. Preferably, the codon-optimised gag-pol genes of the invention may comprise or consist of a
20 nucleic acid sequence having at least 90%, more preferably at least 95%, even more preferably at least 98%, or more sequence identity to SEQ ID NO: 1. Accordingly, the invention provides a nucleic acid comprising codon-optimised gag-pol genes, said nucleic acid having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to SEQ ID NO: 1, preferably at least 90%, more preferably at least 95%, even more preferably at least
25 98%, or more sequence identity to SEQ ID NO: 1. In a particularly preferred embodiment, the invention provides a nucleic acid which comprises or consists of the nucleic acid sequence of SEQ ID NO: 1. The codon-optimised gag-pol genes (e.g. SIV gag-pol genes) of the invention are typically operably linked to a promoter to facilitate expression of the gag-pol proteins. Any suitable promoter may be used, including those described herein in the context of promoters for the transgene.
30 Preferably, the promoter is a CAG promoter, as used on the exemplified pGM691 plasmid. An exemplary CAG promoter is set out in SEQ ID NO: 29. The codon-optimised gag-pol genes of SEQ ID NO: 1 comprise a translational slip, and so do not form a single conventional open reading frame.

The invention also provides plasmids comprising the codon-optimised SIV gag-pol genes of the invention, i.e. pDNA2a comprising the codon-optimised SIV gag-pol genes of the invention. These

plasmids are typically suitable for use in the methods of the invention. The (pDNA2a) plasmid of the invention may comprise or consist of a nucleic acid sequence of SEQ ID NO: 5 (pGM691), or a variant thereof (as defined herein). In particular, the (pDNA2a) plasmid of the invention may comprise or consist of a nucleic acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 5 96%, at least 97%, at least 98%, at least 99% or more sequence identity to SEQ ID NO: 5. Preferably, the (pDNA2a) plasmid of the invention may comprise or consist of a nucleic acid sequence having at least 90%, more preferably at least 95%, even more preferably at least 98%, or more sequence identity to SEQ ID NO: 5. Accordingly, the invention provides a plasmid comprising codon-optimised SIV gag-pol genes of the invention (as defined herein), particularly, a nucleic acid sequence comprising or 10 consisting of SEQ ID NO: 1, or a variant thereof (as defined herein). Said plasmid may comprise or consist of a nucleic acid sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to SEQ ID NO: 5, preferably at least 90%, more preferably at least 95%, even more preferably at least 98%, or more sequence identity to SEQ ID NO: 5. In a particularly preferred embodiment, the invention provides a plasmid which 15 comprises or consists of the nucleic acid sequence of SEQ ID NO: 5. In the plasmid of SEQ ID NO: 5 (or variants thereof): (i) the codon-optimised gag-pol genes of SEQ ID NO: 1 comprise a translational slip, and so do not form a single conventional open reading frame; and (ii) the codon-optimised gag-pol genes of SEQ ID NO: 1 are operably linked to a CAG promoter (e.g. as exemplified herein).

The codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof) and 20 plasmids comprising said genes or nucleic acids are advantageous in the production of retroviral/lentiviral (e.g. SIV) vectors using methods of the invention, as they allow for the production of high titre F/HN retroviral/lentiviral (e.g. SIV) vectors. Typically said codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof) and plasmids comprising said genes or nucleic acids can be used to produce a titre of retroviral/lentiviral (e.g. SIV) vector that is at least equivalent to the 25 titre of retroviral/lentiviral (e.g. SIV) vector produced by a corresponding method which does not use codon-optimised gag-pol genes, as described herein.

Preferably, the codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof) and plasmids comprising said genes or nucleic acids allow for the production of a titre of retroviral/lentiviral (e.g. SIV) vector that is increased compared with the titre of retroviral/lentiviral 30 (e.g. SIV) vector produced by a corresponding method which does not use codon-optimised gag-pol genes, as described herein.

The invention also provides host cells comprising (i) a retroviral/lentiviral (e.g. SIV) vector of the invention, (ii) codon-optimised gag-pol genes (or a nucleic acid comprising or consisting thereof) of the invention; and/or (iii) a plasmid comprising said genes or nucleic acid; or any combination

thereof. Typically a host cell is a mammalian cell, particularly a human cell or cell line. Non-limiting examples of host cells include HEK293 cells (such as HEK293F or HEK293T cells) and 293T/17 cells. Commercial cell lines suitable for the production of virus are also readily available (as described herein).

5 The invention also provides a retroviral/lentiviral (e.g. SIV) vector obtainable by a method of the invention, or using codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention.

Typically the retroviral/lentiviral (e.g. SIV) vector obtainable by a method of the invention, or using codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid
10 comprising said genes or nucleic acids, or host cell of the invention is produced at a high-titre. Titre may be measured in terms of transducing units, as defined here. As described herein, the methods of the invention typically produce retroviral/lentiviral (e.g. SIV) vector at equivalent or higher titres than corresponding methods which do not use codon-optimised gag-pol genes. Accordingly, the retroviral/lentiviral (e.g. SIV) vector obtainable by a method of the invention, or using codon-
15 optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention may optionally be at a titre of at least about 2.5×10^6 TU/mL, at least about 3.0×10^6 TU/mL, at least about 3.1×10^6 TU/mL, at least about 3.2×10^6 TU/mL, at least about 3.3×10^6 TU/mL, at least about 3.4×10^6 TU/mL, at least about 3.5×10^6 TU/mL, at least about 3.6×10^6 TU/mL, at least about 3.7×10^6 TU/mL, at least about 3.8×10^6 TU/mL, at least about 3.9×10^6
20 TU/mL, at least about 4.0×10^6 TU/mL or more. Preferably the retroviral/lentiviral (e.g. SIV) vector is produced at a titre of at least about 3.0×10^6 TU/mL, or at least about 3.5×10^6 TU/mL.

The production of high-titre retroviral/lentiviral (e.g. SIV) vectors may impart other desirable properties on the resulting vector products. For example, without being bound by theory, it is believed that production at high titres without the need for intense concentration by methods such as TFF
25 results in a higher quality vector product than retroviral/lentiviral (e.g. SIV) vectors produced by corresponding methods without the use of codon-optimised gag-pol genes (and optionally a modified vector genome plasmid), because the vectors are exposed to less shear forces which can damage the viral particles and their RNA cargo.

The invention also provides a method of increasing retroviral/lentiviral (e.g. SIV) vector titre
30 comprising the use of codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention. Said method of increasing retroviral/lentiviral (e.g. SIV) vector titre according to the invention may increase titre by at least 1.5-fold, at least 2-fold, or at least 2.5-fold or more compared with a corresponding method which uses non-codon-optimised versions of the gag-pol genes (or nucleic acids comprising or

consisting thereof), or plasmids or host cells comprising said non-codon optimised genes or nucleic acids. Alternatively, a method of increasing retroviral/lentiviral (e.g. SIV) titre according to the invention may increase titre by at least about 25%, at least about 50%, at least about 100%, at least about 150%, at least about 200% or more compared with a corresponding method which uses non-codon-optimised versions of the gag-pol genes (or nucleic acids comprising or consisting thereof), or plasmids or host cells comprising said non-codon optimised genes or nucleic acids. Preferably, a method of increasing retroviral/lentiviral (e.g. SIV) titre according to the invention may increase titre by (a) by at least 1.5-fold or at least 2-fold; and/or (b) by at least about 25%, more preferably at least about 50%, even more preferably at least about 100%. Typically the corresponding method is identical to the method of the invention except for the use of codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention. All the disclosure herein in relation to method of producing a retroviral/lentiviral (e.g. SIV) vector applies equally and without reservation to the methods of increasing retroviral/lentiviral (e.g. SIV) titre of the invention.

The invention also provides the use of codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention to increase the titre of a retroviral/lentiviral (e.g. SIV) vector. Said use may increase retroviral/lentiviral (e.g. SIV) vector titre by at least 1.5-fold, at least 2-fold, or at least 2.5-fold or more compared with the use of a corresponding non-codon-optimised version of the gag-pol genes (or nucleic acids comprising or consisting thereof), or plasmids or host cells comprising said non-codon optimised genes or nucleic acids. Alternatively, said use may increase retroviral/lentiviral (e.g. SIV) titre by at least about 25%, at least about 50%, at least about 100%, at least about 150%, at least about 200% or more compared with the use of a corresponding non-codon-optimised version of the gag-pol genes (or nucleic acids comprising or consisting thereof), or plasmids or host cells comprising said non-codon optimised genes or nucleic acids. Preferably, said use increases retroviral/lentiviral (e.g. SIV) titre by (a) by at least 1.5-fold or at least 2-fold; and/or (b) at least about 25%, more preferably at least about 50%, even more preferably at least about 100%. Typically the corresponding use is identical to the method of the invention except for the use of codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention. All the disclosure herein in relation to method of producing a retroviral/lentiviral (e.g. SIV) vector applies equally and without reservation to the use of codon-optimised gag-pol genes (or nucleic acids comprising or consisting thereof), a plasmid comprising said genes or nucleic acids, or host cell of the invention to increase the titre of a retroviral/lentiviral (e.g. SIV) vector according to the invention. The use of codon-optimised gag-pol genes in combination with a modified vector genome

plasmid (with reduced viral ORFs) may provide a further advantage, in terms of safety and/or vector titre. Thus, the increased vector yields as described herein may be achieved using codon-optimised gag-pol genes alone, or in combination with a modified vector genome plasmid. Any and all disclosure herein in relation to increased vector titre in the context of method using codon-optimised gag-pol
5 genes applies equally and without reservation to methods using codon-optimised gag-pol genes in combination with a modified vector genome plasmid of the invention, and to vectors produced by such methods.

Therapeutic Indications

10 The retroviral/lentiviral (e.g. SIV) vectors of the present invention enable higher and sustained gene expression through efficient gene transfer. The F/HN-pseudotyped retroviral/lentiviral (e.g. SIV) vectors of the invention are capable of: (i) airway transduction without disruption of epithelial integrity; (ii) persistent gene expression; (iii) lack of chronic toxicity; and (iv) efficient repeat administration. Long term/persistent stable gene expression, preferably at a therapeutically-effective
15 level, may be achieved using repeat doses of a vector of the present invention. Alternatively, a single dose may be used to achieve the desired long-term expression.

Thus, advantageously, the retroviral/lentiviral (e.g. SIV) vectors of the present invention can be used in gene therapy. By way of example, the efficient airway cell uptake properties of the retroviral/lentiviral (e.g. SIV) vectors of the invention make them highly suitable for treating
20 respiratory tract diseases. The retroviral/lentiviral (e.g. SIV) vectors of the invention can also be used in methods of gene therapy to promote secretion of therapeutic proteins. By way of further example, the invention provides secretion of therapeutic proteins into the lumen of the respiratory tract or the circulatory system. Thus, administration of a retroviral/lentiviral (e.g. SIV) vector of the invention and its uptake by airway cells may enable the use of the lungs (or nose or airways) as a “factory” to produce
25 a therapeutic protein that is then secreted and enters the general circulation at therapeutic levels, where it can travel to cells/tissues of interest to elicit a therapeutic effect. In contrast to intracellular or membrane proteins, the production of such secreted proteins does not rely on specific disease target cells being transduced, which is a significant advantage and achieves high levels of protein expression. Thus, other diseases which are not respiratory tract diseases, such as cardiovascular
30 diseases and blood disorders, particularly blood clotting deficiencies, can also be treated by the retroviral/lentiviral (e.g. SIV) vectors of the present invention.

Retroviral/lentiviral (e.g. SIV) vectors of the invention can effectively treat a disease by providing a transgene for the correction of the disease. For example, inserting a functional copy of the *CFTR* gene to ameliorate or prevent lung disease in CF patients, independent of the underlying

mutation. Accordingly, retroviral/lentiviral (e.g. SIV) vectors of the invention may be used to treat cystic fibrosis (CF), typically by gene therapy with a CFTR transgene as described herein.

As another example, retroviral/lentiviral (e.g. SIV) vectors of the invention may be used to treat Alpha-1 Antitrypsin (A1AT) deficiency, typically by gene therapy with a A1AT transgene as described herein. A1AT is a secreted anti-protease that is produced mainly in the liver and then trafficked to the lung, with smaller amounts also being produced in the lung itself. The main function of A1AT is to bind and neutralise/inhibit neutrophil elastase. Gene therapy with A1AT according to the present invention is relevant to A1AT deficient patient, as well as in other lung diseases such as CF or chronic obstructive pulmonary disease (COPD), and offers the opportunity to overcome some of the problems encountered by conventional enzyme replacement therapy (in which A1AT isolated from human blood and administered intravenously every week), providing stable, long-lasting expression in the target tissue (lung/nasal epithelium), ease of administration and unlimited availability.

Transduction with a retroviral/lentiviral (e.g. SIV) vector of the invention may lead to secretion of the recombinant protein into the lumen of the lung as well as into the circulation. One benefit of this is that the therapeutic protein reaches the interstitium. A1AT gene therapy may therefore also be beneficial in other disease indications, non-limiting examples of which include type 1 and type 2 diabetes, acute myocardial infarction, ischemic heart disease, rheumatoid arthritis, inflammatory bowel disease, transplant rejection, graft versus host (GvH) disease, multiple sclerosis, liver disease, cirrhosis, vasculitides and infections, such as bacterial and/or viral infections.

A1AT has numerous other anti-inflammatory and tissue-protective effects, for example in pre-clinical models of diabetes, graft versus host disease and inflammatory bowel disease. The production of A1AT in the lung and/or nose following transduction according to the present invention may, therefore, be more widely applicable, including to these indications.

Other examples of diseases that may be treated with gene therapy of a secreted protein according to the present invention include cardiovascular diseases and blood disorders, particularly blood clotting deficiencies such as haemophilia (A, B or C), von Willebrand disease and Factor VII deficiency.

Other examples of diseases or disorders to be treated include Primary Ciliary Dyskinesia (PCD), acute lung injury, Surfactant Protein B (SFTB) deficiency, Pulmonary Alveolar Proteinosis (PAP), Chronic Obstructive Pulmonary Disease (COPD) and/or inflammatory, infectious, immune or metabolic conditions, such as lysosomal storage diseases.

Accordingly, the invention provides a method of treating a disease, the method comprising administering a retroviral/lentiviral (e.g. SIV) vector of the invention to a subject. Typically the retroviral/lentiviral (e.g. SIV) vector is produced using a method of the present invention. Any disease

described herein may be treated according to the invention. In particular, the invention provides a method of treating a lung disease using a retroviral/lentiviral (e.g. SIV) vector of the invention. The disease to be treated may be a chronic disease. Preferably, a method of treating CF is provided.

The invention also provides a retroviral/lentiviral (e.g. SIV) vector as described herein for use
5 in a method of treating a disease. Typically the retroviral/lentiviral (e.g. SIV) vector is produced using a method of the present invention. Any disease described herein may be treated according to the invention. In particular, the invention provides a retroviral/lentiviral (e.g. SIV) vector of the invention for use in a method of treating a lung disease. The disease to be treated may be a chronic disease. Preferably, a retroviral/lentiviral (e.g. SIV) vector for use in treating CF is provided.

10 The invention also provides the use of a retroviral/lentiviral (e.g. SIV) vector as described herein in the manufacture of a medicament for use in a method of treating a disease. Typically the retroviral/lentiviral (e.g. SIV) vector is produced using a method of the present invention. Any disease described herein may be treated according to the invention. In particular, the invention provides the use of a retroviral/lentiviral (e.g. SIV) vector of the invention for the manufacture of a medicament for
15 use in a method of treating a lung disease. The disease to be treated may be a chronic disease. Preferably, the use of a retroviral/lentiviral (e.g. SIV) vector in the manufacture of a medicament for use in a method of treating CF is provided.

Formulation and administration

20 The retroviral/lentiviral (e.g. SIV) vectors of the invention may be administered in any dosage appropriate for achieving the desired therapeutic effect. Appropriate dosages may be determined by a clinician or other medical practitioner using standard techniques and within the normal course of their work. Non-limiting examples of suitable dosages include 1×10^8 transduction units (TU), 1×10^9 TU, 1×10^{10} TU, 1×10^{11} TU or more.

25 The invention also provides compositions comprising the retroviral/lentiviral (e.g. SIV) vectors described above, and a pharmaceutically-acceptable carrier. Non-limiting examples of pharmaceutically acceptable carriers include water, saline, and phosphate-buffered saline. In some embodiments, however, the composition is in lyophilized form, in which case it may include a stabilizer, such as bovine serum albumin (BSA). In some embodiments, it may be desirable to
30 formulate the composition with a preservative, such as thiomersal or sodium azide, to facilitate long-term storage.

The retroviral/lentiviral (e.g. SIV) vectors of the invention may be administered by any appropriate route. It may be desired to direct the compositions of the present invention (as described above) to the respiratory system of a subject. Efficient transmission of a therapeutic/prophylactic

composition or medicament to the site of infection in the respiratory tract may be achieved by oral or intra-nasal administration, for example, as aerosols (e.g. nasal sprays), or by catheters. Typically the retroviral/lentiviral (e.g. SIV) vectors of the invention are stable in clinically relevant nebulisers, inhalers (including metered dose inhalers), catheters and aerosols, etc.

5 In some embodiments the nose is a preferred production site for a therapeutic protein using a retroviral/lentiviral (e.g. SIV) vector of the invention for at least one of the following reasons: (i) extracellular barriers such as inflammatory cells and sputum are less pronounced in the nose; (ii) ease of vector administration; (iii) smaller quantities of vector required; and (iv) ethical considerations. Thus, transduction of nasal epithelial cells with a retroviral/lentiviral (e.g. SIV) vector of the invention
10 may result in efficient (high-level) and long-lasting expression of the therapeutic transgene of interest. Accordingly, nasal administration of a retroviral/lentiviral (e.g. SIV) vector of the invention may be preferred.

Formulations for intra-nasal administration may be in the form of nasal droplets or a nasal spray. An intra-nasal formulation may comprise droplets having approximate diameters in the range
15 of 100-5000 μm , such as 500-4000 μm , 1000-3000 μm or 100-1000 μm . Alternatively, in terms of volume, the droplets may be in the range of about 0.001-100 μl , such as 0.1-50 μl or 1.0-25 μl , or such as 0.001-1 μl .

The aerosol formulation may take the form of a powder, suspension or solution. The size of aerosol particles is relevant to the delivery capability of an aerosol. Smaller particles may travel further
20 down the respiratory airway towards the alveoli than would larger particles. In one embodiment, the aerosol particles have a diameter distribution to facilitate delivery along the entire length of the bronchi, bronchioles, and alveoli. Alternatively, the particle size distribution may be selected to target a particular section of the respiratory airway, for example the alveoli. In the case of aerosol delivery of the medicament, the particles may have diameters in the approximate range of 0.1-50 μm ,
25 preferably 1-25 μm , more preferably 1-5 μm .

Aerosol particles may be for delivery using a nebulizer (e.g. via the mouth) or nasal spray. An aerosol formulation may optionally contain a propellant and/or surfactant.

The formulation of pharmaceutical aerosols is routine to those skilled in the art, see for example, Sciarra, J. in Remington's Pharmaceutical Sciences (*supra*). The agents may be formulated as
30 solution aerosols, dispersion or suspension aerosols of dry powders, emulsions or semisolid preparations. The aerosol may be delivered using any propellant system known to those skilled in the art. The aerosols may be applied to the upper respiratory tract, for example by nasal inhalation, or to the lower respiratory tract or to both. The part of the lung that the medicament is delivered to may be determined by the disorder. Compositions comprising a vector of the invention, in particular where

intranasal delivery is to be used, may comprise a humectant. This may help reduce or prevent drying of the mucus membrane and to prevent irritation of the membranes. Suitable humectants include, for instance, sorbitol, mineral oil, vegetable oil and glycerol; soothing agents; membrane conditioners; sweeteners; and combinations thereof. The compositions may comprise a surfactant. Suitable surfactants include non-ionic, anionic and cationic surfactants. Examples of surfactants that may be used include, for example, polyoxyethylene derivatives of fatty acid partial esters of sorbitol anhydrides, such as for example, Tween 80, Polyoxyl 40 Stearate, Polyoxy ethylene 50 Stearate, fusieates, bile salts and Octoxynol.

In some cases after an initial administration a subsequent administration of a retroviral/lentiviral (e.g. SIV) vector may be performed. The administration may, for instance, be at least a week, two weeks, a month, two months, six months, a year or more after the initial administration. In some instances, retroviral/lentiviral (e.g. SIV) vector of the invention may be administered at least once a week, once a fortnight, once a month, every two months, every six months, annually or at longer intervals. Preferably, administration is every six months, more preferably annually. The retroviral/lentiviral (e.g. SIV) vectors may, for instance, be administered at intervals dictated by when the effects of the previous administration are decreasing.

Any two or more retroviral/lentiviral (e.g. SIV) vectors of the invention may be administered separately, sequentially or simultaneously. Thus two retroviral/lentiviral (e.g. SIV) vectors or more retroviral/lentiviral (e.g. SIV) vectors, where at least one retroviral/lentiviral (e.g. SIV) vectors is a retroviral/lentiviral (e.g. SIV) vector of the invention, may be administered separately, simultaneously or sequentially and in particular two or more retroviral/lentiviral (e.g. SIV) vectors of the invention may be administered in such a manner. The two may be administered in the same or different compositions. In a preferred instance, the two retroviral/lentiviral (e.g. SIV) vectors may be delivered in the same composition.

25

SEQUENCE HOMOLOGY

Any of a variety of sequence alignment methods can be used to determine percent identity, including, without limitation, global methods, local methods and hybrid methods, such as, e.g., segment approach methods. Protocols to determine percent identity are routine procedures within the scope of one skilled in the art. Global methods align sequences from the beginning to the end of the molecule and determine the best alignment by adding up scores of individual residue pairs and by imposing gap penalties. Non-limiting methods include, e.g., CLUSTAL W, see, e.g., Julie D. Thompson et al., CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position- Specific Gap Penalties and Weight Matrix Choice, 22(22) Nucleic Acids

- Research 4673-4680 (1994); and iterative refinement, see, e.g., Osamu Gotoh, Significant Improvement in Accuracy of Multiple Protein. Sequence Alignments by Iterative Refinement as Assessed by Reference to Structural Alignments, 264(4) J. Mol. Biol. 823-838 (1996). Local methods align sequences by identifying one or more conserved motifs shared by all of the input sequences.
- 5 Non-limiting methods include, e.g., Match-box, see, e.g., Eric Depiereux and Ernest Feytmans, Match-Box: A Fundamentally New Algorithm for the Simultaneous Alignment of Several Protein Sequences, 8(5) CABIOS 501 -509 (1992); Gibbs sampling, see, e.g., C. E. Lawrence et al., Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment, 262(5131) Science 208-214 (1993); Align-M, see, e.g., Ivo Van Walle et al., Align-M - A New Algorithm for Multiple Alignment of
- 10 Highly Divergent Sequences, 20(9) Bioinformatics:1428-1435 (2004).

Thus, percent sequence identity is determined by conventional methods. See, for example, Altschul et al., Bull. Math. Bio. 48: 603-16, 1986 and Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915-19, 1992. Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "blosum 62" scoring matrix

15 of Henikoff and Henikoff (ibid.) as shown below (amino acids are indicated by the standard one-letter codes).

The "percent sequence identity" between two or more nucleic acid or amino acid sequences is a function of the number of identical positions shared by the sequences. Thus, % identity may be calculated as the number of identical nucleotides / amino acids divided by the total number of

20 nucleotides / amino acids, multiplied by 100. Calculations of % sequence identity may also take into account the number of gaps, and the length of each gap that needs to be introduced to optimize alignment of two or more sequences. Sequence comparisons and the determination of percent identity between two or more sequences can be carried out using specific mathematical algorithms, such as BLAST, which will be familiar to a skilled person.

25

ALIGNMENT SCORES FOR DETERMINING SEQUENCE IDENTITY

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
30 N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													

G 0 -2 0 -1 -3 -2 -2 6
 H -2 0 1 -1 -3 0 0 -2 8
 I -1 -3 -3 -3 -1 -3 -3 -4 -3 4
 L -1 -2 -3 -4 -1 -2 -3 -4 -3 2 4
 5 K -1 2 0 -1 -3 1 1 -2 -1 -3 -2 5
 M -1 -1 -2 -3 -1 0 -2 -3 -2 1 2 -1 5
 F -2 -3 -3 -3 -2 -3 -3 -3 -1 0 0 -3 0 6
 P -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4 7
 S 1 -1 1 0 -1 0 0 0 -1 -2 -2 0 -1 -2 -1 4
 10 T 0 -1 0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1 1 5
 W -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1 1 -4 -3 -2 11
 Y -2 -2 -2 -3 -2 -1 -2 -3 2 -1 -1 -2 -1 3 -3 -2 -2 2 7
 V 0 -3 -3 -3 -1 -2 -2 -3 -3 3 1 -2 1 -1 -2 -2 0 -3 -1 4

15 The percent identity is then calculated as:

Total number of identical matches
 _____ x 100
 [length of the longer sequence plus the
 20 number of gaps introduced into the longer
 sequence in order to align the two sequences]

Substantially homologous polypeptides are characterized as having one or more amino acid
 substitutions, deletions or additions. These changes are preferably of a minor nature, that is
 25 conservative amino acid substitutions (as described herein) and other substitutions that do not
 significantly affect the folding or activity of the polypeptide; small deletions, typically of one to about
 30 amino acids; and small amino- or carboxyl-terminal extensions, such as an amino-terminal
 methionine residue, a small linker peptide of up to about 20-25 residues, or an affinity tag.

In addition to the 20 standard amino acids, non-standard amino acids (such as 4-
 30 hydroxyproline, 6-N-methyl lysine, 2-aminoisobutyric acid, isovaline and α -methyl serine) may be
 substituted for amino acid residues of the polypeptides of the present invention. A limited number of
 non-conservative amino acids, amino acids that are not encoded by the genetic code, and unnatural
 amino acids may be substituted for polypeptide amino acid residues. The polypeptides of the present
 invention can also comprise non-naturally occurring amino acid residues.

Non-naturally occurring amino acids include, without limitation, trans-3-methylproline, 2,4-methano-proline, cis-4-hydroxyproline, trans-4-hydroxy-proline, N-methylglycine, allo-threonine, methyl-threonine, hydroxy-ethylcysteine, hydroxyethylhomo-cysteine, nitro-glutamine, homoglutamine, pipecolic acid, tert-leucine, norvaline, 2-azaphenylalanine, 3-azaphenyl-alanine, 4-azaphenyl-alanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an in vitro system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell free system comprising an E. coli S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., J. Am. Chem. Soc. 113:2722, 1991; Ellman et al., Methods Enzymol. 202:301, 1991; Chung et al., Science 259:806-9, 1993; and Chung et al., Proc. Natl. Acad. Sci. USA 90:10145-9, 1993). In a second method, translation is carried out in *Xenopus* oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., J. Biol. Chem. 271:19991-8, 1996). Within a third method, E. coli cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the polypeptide in place of its natural counterpart. See, Koide et al., Biochem. 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by in vitro chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, Protein Sci. 2:395-403, 1993).

A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, non-naturally occurring amino acids, and unnatural amino acids may be substituted for amino acid residues of polypeptides of the present invention.

Essential amino acids in the polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, Science 244: 1081-5, 1989). Sites of biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., Science 255:306-12, 1992; Smith et al., J. Mol. Biol. 224:899-904, 1992; Wlodaver et al., FEBS Lett. 309:59-64, 1992. The identities of

essential amino acids can also be inferred from analysis of homologies with related components (e.g. the translocation or protease components) of the polypeptides of the present invention.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (Science 241:53-7, 1988) or Bowie and Sauer (Proc. Natl. Acad. Sci. USA 86:2152-6, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., Biochem. 30:10832-7, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., Gene 46:145, 1986; Ner et al., DNA 7:127, 1988).

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (Science 241:53-7, 1988) or Bowie and Sauer (Proc. Natl. Acad. Sci. USA 86:2152-6, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., Biochem. 30:10832-7, 1991; Ladner et al., U.S. Patent No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., Gene 46:145, 1986; Ner et al., DNA 7:127, 1988).

EXAMPLES

The invention is now described with reference to the Examples below. These are not limiting on the scope of the invention, and a person skilled in the art would be appreciate that suitable equivalents could be used within the scope of the present invention. Thus, the Examples may be considered component parts of the invention, and the individual aspects described therein may be considered as disclosed independently, or in any combination.

Example 1 – Plasmid pGM691 construction

A comparison of the vector genome plasmid (pDNA1) of pGM326 with the GagPol plasmid (pDNA2a) of pGM297 was carried out. As shown in Figure 5A, there is significant homology between the partial gagpol nucleotide sequence in pGM326 and the non-codon optimised gagpol sequence of pGM297.

A modified pDNA2a plasmid was designed to (i) reduce the homology between the partial gagpol nucleotide sequence in pGM326 and the non-codon optimised gagpol sequence of pGM297; (ii) to codon-optimize the *gagpol* genes for increased gagpol protein expression; (iii) to reduce the theoretical risk of generating replication-competent lentivirus (RCL) during manufacture or clinical use; and (iv) to eliminate gagpol expression dependency on Rev. A comparison of pGM297 with the modified pDNA2a (pGM691) is shown in Figures 5B-5D, with the changes annotated.

pGM691 was created by digesting pGM297 with the restriction enzymes XhoI, EcoRV and BglII to yield DNA fragments of 4583bp, 3662bp and 1641bp. The 4583bp fragment, containing the plasmid origin of replication and CBA promoter intron was purified and retained. The plasmid pGM693 was manufactured by GeneArt/LifeTechnologies via DNA synthesis. pGM693 was designed by the inventors to include a 4481bp XhoI to BglII DNA fragment that included the codon optimised GagPol sequence ultimately found in pGM691. pGM693 was digested with XhoI and BglII to yield DNA fragments of 4481bp, 1236bp and 1048bp. The 4481bp fragment, containing the codon optimised GagPol sequence was purified and retained (see Figure 5E). The two retained DNA fragments were ligated with DNA ligase and the resulting mixture of ligated DNA was transformed into *E. coli* Stbl3 cells; cells containing plasmids capable of replication were selected by resistance to kanamycin. Well-isolated individual colonies of kanamycin resistant, transformed Stbl3 cells were selected and expanded. DNA restriction analysis of the resultant clones identified a number of clones with the expected DNA structure; one was reserved and termed pGM691.

Example 2 – Production of rSIV.F/HN vector hCEF-CFTR

The vector genome pGM326, which incorporates a CFTR transgene under the transcriptional control of the hCEF promoter was used in two design of experiments (DoE) studies to evaluate the production yields provided by using either pGM297 GagPol or pGM691 coGagPol.

In each DoE study a wide range of conditions was employed that included low, centre and high concentrations of each of the components used:

Function	Code	Low	Centre	High
Genome	pGM326	0.2	1.1	2
(co)GagPol	pGM297 or GM691	0.1	0.55	1
Rev	pGM299	0.1	0.55	1
F	pGM301	0.1	0.55	1
HN	pGM303	0.1	0.55	1
Transfection Reagent	Lipofectamine 2000	4	7	10

The units for transfection reagent was $\mu\text{L/mL}$, for all other reagents it was $\mu\text{g/mL}$.

A 3-level fractional factorial design was employed with duplicate vector stocks prepared for the majority of conditions and six replicate centre points. Overall, 31 vector stocks were prepared using otherwise identical conditions for pGM297 GagPol and pGM691 coGagPol.

5 The integrating transducing unit titre (TU/mL), as determined by the detection of the ratio of vector specific and genome specific DNA sequences in transduced cells via quantitative PCR following transduction of 293T cells with dilutions of the vector stocks was plotted in Figure 6A (replicate vector stocks represented as dots, the line indicates otherwise identical conditions).

10 Following on from the DOE experiments, vector genome pGM326, which incorporates a CFTR transgene under the transcriptional control of the hCEF promoter was used to prepare rSIV.F/HN vector stocks in triplicate using either pGM297 GagPol or pGM691 coGagPol as indicated.

For all preparations, Rev, F and HN were provided by pGM299, pGM301 and pGM303 respectively. The DNA mass ratio of vector genome:GagPol:Rev:F:HN used was 20:9:6:6:6 in all cases. For conditions A and B, the total DNA levels used were 2.2µg/mL and 1.8µg/mL respectively. For
15 conditions A and B, the total Lipofectamine 2000 levels used were 7µL/mL and 8µL/mL respectively.

The integrating transducing unit titre (TU/mL), as determined by the ratio of vector specific to genome specific DNA sequences in transduced cells via quantitative PCR following transduction of 293T cells with dilutions of the vector stocks, is plotted (individual vector stocks represented as dots, the line indicates the group median).

20 Vector yields with the coGagPol as provided by pGM691 was observed to be ~2.3-fold higher under Condition A and ~1.5-fold higher under Condition B (Figure 6B). Thus, use of pGM691 as pDNA2a observably increased SIV viral titre, independent of other culture conditions used. This is surprising, because there are multiple independent published studies which report that codon-optimisation of the gagpol genes is associated with a decrease in lentiviral titre.

25

Example 3 – Production of rSIV.F/HN CMV-EGFP

To investigate whether or not the ability of codon-optimised gagpol to maintain or increase vector titre was limited to the specific rSIV.F/HN construct (rSIV.F/HN hCEF-CFTR), experiments were conducted using plasmids to produce a different transgene operably linked to a different promoter.

30 HEK293T, Freestyle 293F (Life Technologies, Paisley, UK) and 293T/17 cells (CRL-11268; ATCC, Manassas, VA) were maintained in Dulbecco's minimal Eagle's medium (Invitrogen, Carlsbad, CA) containing 10% fetal bovine serum and supplemented with penicillin (100 U/ml) and streptomycin (100 µg/ml) or Freestyle™ 293 Expression Medium (Life Technologies).

SeV-F/HN-pseudotyped SIV vector was produced by transfecting HEK293T or 293T/17 cells cultured in FreeStyle™ 293 Expression Medium with a mixture of five plasmids with the following characteristics: pDNA1 (pGM311; which incorporates an EGFP transgene under the transcriptional control of the CMV promoter) encodes the lentiviral vector mRNA; pDNA2a (pGM691; Figure 2C) encodes SIV Gag and Pol proteins; pDNA2b (pGM299; Figure 2D) encodes SIV Rev proteins; pDNA3a (pGM301; Figure 2E) encodes the Sendai virus-derived Fct4 protein [Kobayashi et al., 2003 J. Virol. 77:2607]; and pDNA3b (pGM303; Figure 2F) encodes the Sendai virus-derived SIVct+HN [Kobayashi et al., 2003 J. Virol. 77:2607] complexed with PEIpro (Polyplus, Illkirch, France). Cell culture media was supplemented at 12-24 post-transfection with sodium butyrate. Sodium butyrate stimulates vector production via inhibiting histone deacetylase resulting in increasing expression of the SIV and Sendai virus fusion protein components encoded by the five plasmids. Cell culture media was supplemented at 44-52 hours and/or 68-76 hours post-transfection with 5 units/mL Benzonase Nuclease (Merck Millipore, Nottingham, UK). The culture supernatant containing the SIV vector was harvested 68-76.5 hours after transfection, and clarified by filtration through a 0.45 µm membrane. The SIV vector is treated by digestion with TrypLE Select™. Subsequently, SIV vector was further purified and concentrated by anion-exchange chromatography and tangential flow filtration.

rSIV.F/HN vector stocks in triplicate using either pGM297 GagPol or pGM691 coGagPol as indicated. The DNA mass ratio of vector genome:GagPol:Rev:F:HN used was 20:9:6:6:6 in all cases.

The functional transducing unit titre (FTU/mL), as determined by the detection of EGFP positive cells via flow cytometry following transduction of 293T cells with dilutions of the vector stocks was plotted in Figure 7 (individual vector stocks represented as dots, the line indicates the group median). As for the rSIV.F/HN hCEF-CFTR constructs in Example 2, rSIV.F/HN CMV-EGFP vector yields with the coGagPol as provided by pGM691 were observed to be ~1.6-fold higher than when the non-codon-optimised gagpol of pGM297 was used. This suggests that the ability of codon-optimised gagpol to maintain or increase vector titre was not limited to the specific rSIV.F/HN hCEF-CFTR construct, but rather is a function generally associated with the use of coGagPol.

Example 3 – Reducing the number of intact SIV ORFs within the vector genome plasmid

Additional modifications to one or more of the construction plasmids can further improve the safety of the final vector product, providing a further clinical advantage.

The inventors reviewed sequences of the construction plasmids and identified several regions of concern within the vector genome plasmid pGM326. In particular, the pGM326 partial Gag RRE cPPT hCEF region contains:

- 77 start codons (ATGs);
- 32 ORFs $\geq$ 10 amino acids in length
- 2 large ORFs in the 5' to 3' direction
 - 189 amino acids from the most 5' ATG in vector genome (Gag/RRE fusion),
5 encoding p17 Matrix and part of p24 capsid
 - 250 amino acids from ATG internal to RRE (RRE/cPPT/hCEF fusion)

These are illustrated in Figure 8. The 2 large ORFs (shown in Figure 9) were of particular concern.

10 As such, the inventors designed a modified version of the pGM326 plasmid with a combination of additional modifications intended to reduce the number of intact SIV ORFs (and in particular to remove these 2 large ORFs) for improved safety. The modifications are made to the 2 large ORFs upstream of the hCEF promoter and CFTR transgene (soCFTR2). The changes made were as follows:

- 15 • 6 ATGs Eliminated (3xATG-ATTG, 1xATG-TTG, 2xATG-AAG)
- 1 Stop inserted (TCC-TAAA)
- 1 Restriction site between partial Gag and RRE altered (EcoRI GAATTC – GCCTGCAGG SbfI)

The resulting vector genome plasmid is pGM830 as shown in Figure 2B, with the sequence of
20 SEQ ID NO: 4.

Comparisons of vector titre using either the pGM326 or pGM830 vector genome plasmids in an otherwise identical production protocol demonstrated that the use of pGM830 gave a comparable titre to pGM326 using both HEK293T and A549 cells (see Figure 10), indicating that an improved safety profile could be achieved without adversely affecting titre.

25

Example 4 - Combination of coGagPol and a modified vector genome plasmid maintains, or even increases vector titre

The experiments reported in Example 2 surprisingly demonstrated that, rather than the expected decrease in yield, generation of SIV.F/HN hCEF-CFTR using coGagPol trended to maintain or
30 even increase vector titre. The experiments reported in Example 3 demonstrated that a further improvement to the safety profile of the vector could be achieved by modifying the vector genome plasmid, without adversely affecting the vector titre.

Following on from this, additional experiments were carried out in which the use of coGagPol was combined with the use of the pGM830 vector genome plasmid, to investigate whether these two safety-related modifications could be combined and vector titre maintained.

As illustrated in Figure 11, the inventors surprisingly found that not only could the use of coGagPol be combined with the use of a modified vector genome plasmid (pGM830), but that this combination gave an observable trend to increase vector titre.

This suggests not only can vectors with further improved safety profiles be obtained by combining the use of coGagPol with a modified vector genome plasmid, but that surprisingly this can be achieved whilst maintaining or even increasing rSIV.F/HN hCEF-transgene titre.

10

SEQUENCE INFORMATION

Key to Sequences

- 15 SEQ ID NO: 1 codon-optimised SIV gal-pol nucleic acid sequence
- SEQ ID NO: 2 wild-type SIV gag-pol nucleic acid sequence
- SEQ ID NO: 3 Plasmid as defined in Figure 2A (pDNA1 pGM326)
- SEQ ID NO: 4 Plasmid as defined in Figure 2B (pDNA1 pGM830)
- SEQ ID NO: 5 Plasmid as defined in Figure 2C (pDNA2a pGM691)
- 20 SEQ ID NO: 6 Plasmid as defined in Figure 2D (pDNA2b pGM299)
- SEQ ID NO: 7 Plasmid as defined in Figure 2E (pDNA3a pGM301)
- SEQ ID NO: 8 Plasmid as defined in Figure 2F (pDNA3b pGM303)
- SEQ ID NO: 9 Plasmid as defined in Figure 2G (pDNA2a pGM297)
- SEQ ID NO: 10 Exemplified hCEF promoter
- 25 SEQ ID NO: 11 Exemplified CMV promoter
- SEQ ID NO: 12 Exemplified EF1a promoter
- SEQ ID NO: 13 Exemplified CFTR transgene (*soCFTR2*)
- SEQ ID NO: 14 Exemplified A1AT transgene
- SEQ ID NO: 15 Complementary strand to the exemplified A1AT transgene
- 30 SEQ ID NO: 16 Exemplified A1A1 polypeptide
- SEQ ID NO: 17 Exemplified FVIII transgene (N6)
- SEQ ID NO: 18 Exemplified FVIII transgene (V3)
- SEQ ID NO: 19 Complementary strand to the exemplified FVIII transgene (N6)
- SEQ ID NO: 20 Complementary strand to the exemplified FVIII transgene (V3)

SEQ ID NO: 21 Exemplified FVIII polypeptide (N6)

SEQ ID NO: 22 Exemplified FVIII polypeptide (V3)

SEQ ID NO: 23 Exemplified WPRE component (mWPRE)

SEQ ID NO: 24 F/HN-SIV-hCEF-soA1AT plasmid as defined in Figure 3 (pDNA1 pGM407)

5 SEQ ID NO: 25 F/HN-SIV-CMV-HFVIII-V3 plasmid as defined in Figure 4A (pDNA1 pGM411)

SEQ ID NO: 26 F/HN-SIV-hCEF-HFVIII-V3 plasmid as defined in Figure 4B (pDNA1 pGM413)

SEQ ID NO: 27 F/HN-SIV-CMV-HFVIII-N6-co plasmid as defined in Figure 4C (pDNA1 pGM412)

SEQ ID NO: 28 F/HN-SIV-hCEF-HFVIII-N6-co plasmid as defined in Figure 4D (pDNA1 pGM414)

SEQ ID NO: 29 Exemplary CAG promoter

10

Sequences

SEQ ID NO: 1 codon-optimised SIV gal-pol nucleic acid sequence (from pGM691)

Length: 4391; Molecule Type: DNA; Features Location/Qualifiers: source, 1..4391; mol_type, other DNA; note, codon-optimised SIV gal-pol nucleic acid sequence (from pGM691); organism, synthetic construct

15 ATGGGAGCTGCCACATCTGCCCTGAATAGACGGCAGCTGGACCAGTTCGAGAAGATCAGACTGCGGCCCAACGGC
AAGAAGAAGTACCAGATCAAGCACCTGATCTGGGCCGGCAAAGAGATGGAAAGATTGCGCCTGCACGAGCGGCTG
CTGGAACCGGAGGAAGGCTGCAAGAGAATTATCGAGGTGCTGTACCCTCTGGAACCTACCGGCTCTGAGGGCCTG
20 AAGTCCCTGTTCAATCTCGTGTGCGTGTCTACTGCCTGCACAAAGAACAGAAAGTGAAGGACACCGAAGAGGCC
GTGGCCACAGTTAGACAGCACTGCCACCTGGTGGAAAAAGAGAAGTCCGCCACAGAGACAAGCAGCGGCCAGAAG
AAGAACGACAAGGGAATTGCTGCCCCCTCTGGCGGCAGCCAGAATTTTCTGCTCAGCAGCAGGGAACGCCTGG
GTGCACGTTCCACTGAGCCCTAGAACACTGAATGCCTGGGTCAAAGCCGTGGAAGAGAAGAAGTTTGGCGCCGAG
ATCGTGCCCATGTTCCAGGCTCTGTCTGAGGGCTGCACCCCTTACGACATCAACCAGATGCTGAACGTGCTGGGA
25 GATCACCAGGGCGCTCTGCAGATCGTGAAAGAGATCATCAACGAAGAGGCTGCCAGTGGGACGTGACACATCCA
TTGCCTGCTGGACCTCTGCCAGCCGGACAACCTGAGAGATCCTAGAGGCTCTGATATCGCCGGCACCACCAGCTCT
GTGCAAGAGCAGCTGGAATGGATCTACACCGCCAATCCTAGAGTGGACGTGGGCGCCATCTACAGAAGATGGATC
ATCCTGGGCTGCAGAAATGCGTGAAGATGTACAACCCCGTGTCCGTGCTGGACATCAGACAGGGACCCAAAGAG
CCCTTCAAGGACTACGTGGACCGGTTCTATAAGGCCATTAGAGCCGAGCAGGCCAGCGGCGAAGTGAAGCAGTGG
30 ATGACAGAGAGCCTGCTGATCCAGAACGCCAATCCAGACTGCAAAGTGATCCTGAAAGGCCCTGGGCATGCACCCC
ACACTGGAAGAGATGCTGACAGCCTGTCAAGGCGTTGGCGGCCCTTCTTACAAAGCCAAAGTGATGGCCGAGATG
ATGCAGACCATGCAGAACCAGAACATGGTGCAGCAAGGCGGCCCTAAGAGACAGAGGCCTCCTCTGAGATGCTAC
AACTGCGGCAAGTTCGGCCACATGCAGAGACAGTGTCTGAGCCTAGGAAAACAAAATGCTAAAGTGTGGAAAA
TTGGGACACCTAGCAAAAGACTGCAGGGGACAGGTGAATTTTTTAGGGTATGGACGGTGGATGGGGGCAAAACCG
35 AGAAATTTTCCCGCCGCTACTCTTGAGCGGAACCGAGTGCGCCCTCCTCCACCGAGCGGCACCACCCATACGAC
CCAGCAAAGAAGCTCCTGCAGCAATATGCAGAGAAAGGGAAACAACCTGAGGGAGCAAAAGAGGAATCCACCGGCA
ATGAATCCGGATTGGACCGAGGGATATTCTTTGAACTCCCTCTTTGGAGAAGACCAATAAAGACCGTGTACATCG
AGGGCGTGCCCATCAAGGCTCTGCTGGATACAGGCGCCGACGACACCATCATCAAAGAGAACGACCTGCAGCTGA
GCGGCCCTTGGAGGCCTAAGATCATTTGAGGAATCGGCGGAGGCTGAACGTCAAAGAGTACAACGACCGGGAAG
40 TGAAGATCGAGGACAAGATCCTGAGGGGCACAATCCTGCTGGGCGCCACACCTATCAACATCATCGGCAGAAATC
TGCTGGCCCTTGGCGGCTAGACTGGTTATGGGACAGCTCTCTGAGAAGATCCCCGTGACACCCGTGAAGCTGA
AAGAAGGCGCTAGAGGACCTTGTGTGCGACAGTGGCCTCTGAGCAAAGAGAAGATTGAGGCCCTGCAAGAAATCT
GTAGCCAGCTGGAACAAGAGGGCAAGATCAGCAGAGTTGGCGGCGAGAACGCCTACAATACCCCTATCTTCTGCA
TCAAGAAAAAGACAAGAGCCAGTGGCGGATGCTGTTGGACTTTAGAGAGCTGAACAAGGCTACCCAGGACTTCT
45 TCGAGGTGCAGCTGGGAATTCCTCATCTGCGGCCCTGCGGAAGATGAGACAGATCACAGTGTGGATGTGGGCG
ACGCCTACTACAGCATCCCTCTGGACCCCAACTTCAGAAAGTACACCGCCTTCAATCCCCACCGTGAACAATC

AAGGCCCTGGCATCAGATACCAAGTTCAACTGCCTGCCTCAAGGCTGGAAGGGCAGCCCCACCATTTCAGAAATA
 CCGCCGCCAGCATCCTGGAAGAAATCAAGAGAAACCTGCCTGCTCTGACCATCGTGAGTACATGGACGATCTGT
 GGGTCGGAAGCCAAGAGAATGAGCACACCCACGACAAGCTGGTGGAAACAGCTGAGAACAAAGCTGCAGGCCTGGG
 GCCTCGAAACCCCTGAGAAGAAGGTGCAGAAAGAACCTCCTTACGAGTGGATGGGCTACAAGCTGTGGCCTCACA
 5 AGTGGGAGCTGAGCCGGATTGAGCTCGAAGAGAAGGACGAGTGGACCGTGAACGACATCCAGAAACTCGTGGGCA
 AGCTGAATTGGGCAGCCAGCTGTATCCCGGCTGAGGACCAAGAACATCTGCAAGCTGATCCGGGGAAAGAAGA
 ACCTGCTGGAAGTGGTCACATGGACACCTGAGGCCGAGGCCGAATATGCCGAGAATGCCGAAATCCTGAAAACCG
 AGCAAGAGGGGACCTACTACAAGCCTGGCATTCCAATCAGAGCTGCCGTGCAGAACTGGAAGCGGGCCAGTGGT
 CCTACCAGTTTAAGCAAGAAGGCCAGGTCTGAAAGTGGGCAAGTACACCAAGCAGAAGAACCCACACCAACG
 10 AGCTGAGGACACTGGCTGGCCTGGTCCAGAAAATCTGCAAGAGGCCCTGGTCAATTTGGGGCATCCTGCCTGTTC
 TGGAACTGCCCATTGAGCGGGAAGTGTGGGAACAGTGGTGGGCCGATTACTGGCAAGTGTCTTGGATCCCCGAGT
 GGGACTTCGTGTCTACCCCTCCTCTGCTGAAACTGTGGTACACCCTGACAAAAGAGCCCATTCCTAAAGAGGACG
 TCTACTACGTTGACGGCGCTGCAACCGGAACCTCCAAAGAAGGCAAGGCCGGCTACATCAGCCAGTACGGCAAGC
 AGAGAGTGGAAACCCCTGGAAAACACCACCAACCAGCAGGCCGAGCTGACCGCCATTAAGATGGCCCTGGAAGATA
 15 GCGGCCCAATGTGAACATCGTGACCGACTCTCAGTACGCCATGGGAATCCTGACAGCCCAGCCTACACAGAGCG
 ATAGCCCTCTGGTTGAGCAGATCATTGCCCTGATGATTCAGAAGCAGCAAATCTACCTGCAGTGGGTGCCCGCTC
 ACAAAGGCATCGCGGAAACGAAGAGATCGATAAGCTGGTGTCCAAGGGAATCAGACGGGTGCTGTTCTCTGGAAA
 AGATTGAAGAGGCCCAAGAGGAACACGACGCTACCACAACACTGGAAGAATCTGGCCGACCTACGGACTGTC
 CCCAGATCGTGGCCAAAGAAATCGTGGCTATGTGCCCAAGTGTGAGATCAAGGGCGAACCTGTGCACGGCCAAG
 20 TGGATGCTTCTCCTGGCAGATGGCAGATGGACTGTACCCACCTGGAAGGCAAAGTGGTCACTCGTGGCTGTGCACG
 TGGCCTCCGGCTTTATTGAGGCCAAGTGATCCCCAGAGAGACAGGCAAAGAAAACCGCCAAGTTCCTGCTGAAGA
 TCCTGTCCAGATGGCCCATCACACAGCTGCACACCGACAACGGCCCTAACTTCACATCTCAAGAGGTGGCCGCCA
 TCTGTTGGTGGGGAAGATTGAGCACACAACCGGCATTCCCTACAATCCACAGAGCCAGGGCAGCATCGAGTCCA
 TGAACAAGCAGCTCAAAGAGATTATCGGCAAGATCCGGGACGACTGCCAGTACACAGAAACAGCCGTGCTGATGG
 25 CCTGTACATCCACAACCTCAAGCGGAAAGGCGGCATCGGAGGACAGACATCTGCCGAGAGACTGATCAATATCA
 TCACCCTCAGCTGGAATCCAGCACCTCCAGACCAAGATCCAGAAGATTCTGAACTTCGGGTGTACTACCGCG
 AGGGCAGAGATCCTGTTTGGAAAGGCCAGCACAGCTGATCTGGAAGGCGAAGGTGCCGTGGTGTGAAGGATG
 GCTCTGATCTGAAGGTGGTCCCCAGACGGAAGGCCAAGATTATCAAGGATTACGAGCCCAAACAGCGCTGGGCA
 ATGAAGGCGACGTTGAGGGCACAAGAGGCAGCGACAATTGA
 30

SEQ ID NO: 2 wild-type SIV gag-pol nucleic acid sequence (from pGM297)

Length: 4391; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..4391; mol_type, unassigned DNA; organism, Simian immunodeficiency virus
 ATGGGGGCGGTACCTCAGCACTAAATAGGAGACAATTAGACCAATTTGAGAAAATACGACTTCGCCCCGAACGGA
 35 AAGAAAAAGTACCAAATTAACATTTAATATGGGCAGGCAAGGAGATGGAGCGCTTCGGCTCCATGAGAGGTTG
 TTGGAGACAGAGGAGGGGTGTAAGAATCATAGAAGTCTCTACCCCTAGAACCAACAGGATCGGAGGGCTTA
 AAAAGTCTGTTCATCTTGTGTGCTACTATATTGCTTGCACAAGGAACAGAAAGTGAAGACACAGAGGAAGCA
 GTAGCAACAGTAAGACAACACTGCCATCTAGTGGAAAAAGAAAAAGTGCAACAGAGACATCTAGTGGACAAAAG
 AAAAATGACAAGGGAATAGCAGCGCCACCTGGTGGCAGTCAGAATTTTCCAGCGCAACAACAAGGAAATGCCTGG
 40 GTACATGTACCCCTTGTACCGCGCACCTTAAATGCGTGGGTAAAAGCAGTAGAGGAGAAAAAATTTGAGCAGAA
 ATAGTACCCATGTTTCAAGCCCTATCAGAAGGCTGCACACCTATGACATTAATCAGATGCTTAATGTGCTAGGA
 GATCATCAAGGGGCATTACAAATAGTGAAAGAGATCATTAATGAAGAAGCAGCCAGTGGGATGTAACACACCCA
 CTACCCGCAGGACCCCTACCAGCAGGACAGCTCAGGGACCCCTCGCGGCTCAGATATAGCAGGGACACCAGCTCA
 GTACAAGAACAGTTAGAATGGATCTATACTGCTAACCCCGGGTAGATGTAGGTGCCATCTACCGGAGATGGATT
 45 ATTCTAGGACTTCAAAGTGTGTCAAAATGTACAACCCAGTATCAGTCTTAGACATTAGGCAGGGACCTAAAGAG
 CCCTTCAAGGATTATGTGGACAGATTTTACAAGGCAATTAGAGCAGAACAAGCCTCAGGGGAAGTGAACAAATGG
 ATGACAGAATCATTACTCATTCAAAATGCTAATCCAGATTGTAAGGTATCTCTGAAGGGCTAGGAATGCACCCC
 ACCCTTGAAGAAATGTTAACGGCTTGTGAGGGGTAGGAGGCCCAAGCTACAAAGCAAAAGTAATGGCAGAAATG
 ATGCAGACCATGCAAAATCAAACATGGTGCAGCAGGGAGGTCCAAAAAGACAAAGACCCCACTAAGATGTTAT
 50 AATTGTGGAATAATTTGGCCATATGCAAGACAATGTCCGGAACCAAGGAAAACAAAATGTCTAAAGTGTGGAATA
 TTGGGACACCTAGCAAAAGACTGCAGGGGACAGGTGAATTTTTAGGGTATGGACGGTGGATGGGGGCAAAACCG
 AGAAATTTTCCCGCGCTACTCTTGGAGCGGAACCGAGTGCCTCTCCACCGAGCGGCACCCACCCATACGAC

CCAGCAAAGAAGCTCCTGCAGCAATATGCAGAGAAAGGGAAACAACCTGAGGGAGCAAAAGAGGAATCCACCGGCA
 ATGAATCCGGATTGGACCGAGGGATATTCTTTGAACTCCCTCTTTGGAGAAGACCAATAAAGACAGTGTATATAG
 AAGGGGTCCCCATTAAGGCACTGCTAGACACAGGGGCAGATGACACCATAATTAAGAAAATGATTTACAATTAT
 CAGGTCCATGGAGACCCAAAATTATAGGGGGCATAGGAGGAGGCCTTAATGTAAAGAATATAACGACAGGGAAG
 5 TAAAAATAGAAGATAAAAATTTGAGAGGAACAATATTGTTAGGAGCAACTCCCATTAATATAATAGGTAGAAATT
 TGCTGGCCCCGGCAGGTGCCCGTTAGTAATGGGACAATTATCAGAAAAAATTCCTGTCACACCTGTCAAAATTGA
 AGGAAGGGGCTCGGGGACCTGTGTAAAGACAATGGCCTCTCTCTAAAGAGAAGATTGAAGCTTTACAGGAAATAT
 GTTCCCAATTAGAGCAGGAAGGAAAAATCAGTAGAGTAGGAGGAGAAAAATGCATACAATACCCCAATATTTTGCA
 TAAAGAAGAAGGACAAATCCAGTGGAGGATGCTAGTAGACTTTTAGAGAGTTAAATAAGGCAACCCCAAGATTCT
 10 TTGAAGTGCAATTAGGGATACCCACCCAGCAGGATTAAGAAAGATGAGACAGATAACAGTTTTAGATGTAGGAG
 ACGCCTATTATTCCATACCATTTGGATCCAAATTTTAGGAAATATACTGCTTTTACTATTCCCACAGTGAATAATC
 AGGGACCCGGGATTAGGTATCAATTTCACTGTCTCCCGCAAGGGTGGAAAGGATCTCCTACAATCTTCCAAAATA
 CAGCAGCATCCATTTTGGAGGAGATAAAAAAGAACTTGCCAGCACTAACCATTTGACAATACATGGATGATTTAT
 GGGTAGGTTCTCAAGAAAATGAACACACCCATGACAAATTAGTAGAACAGTTAAGAACAAAATTACAAGCCTGGG
 15 GCTTAGAAACCCAGAAAAGAAGGTGCAAAAAGAACCACCTTATGAGTGGATGGGATACAACTTTGGCCTCACA
 AATGGGAAC TAAGCAGAAATACAACTGGAGGAAAAAGATGAATGGACTGTCAATGACATCCAGAAGTTAGTTGGGA
 AACTAAATTGGGACACAAATTGTATCCAGGTCTTAGGACCAAGAATATATGCAAGTTAATTAGGGAAGAAAA
 ATCTGTTAGAGCTAGTGACTTGGACACCTGAGGCAAGCTGAATATGCAGAAAAATGCAGAGATTCTTAAACAG
 AACAGGAAGGAACCTATTACAAACCAGGAATACCTATTAGGGCAGCAGTACAGAAATTGGAAGGAGGACAGTGGGA
 20 GTTACCAATTCAAACAAGAAGGACAAGTCTTGAAAGTAGGAAAAATACACCAAGCAAAAGAACCCCATACAAATG
 AACTTCGCACATTAGCTGGTTTAGTGCAAGATTTGCAAGAAGCTCTAGTTATTTGGGGGATATTACCAGTTC
 TAGAACTCCCGATAGAAAGAGAGGTATGGGAACAATGGTGGGCGGATTACTGGCAGGTAAGCTGGATTCCCGAAT
 GGGATTTTGTGACGACCCACCTTTGCTCAAACTATGGTACACATTAACAAAAGAACCATAACCCAGGAGGACG
 TTTACTATGTAGATGGAGCATGCAACAGAAATTCAAAAGAAGGAAAAAGCAGGATACATCTCACAATACGGAAAAAC
 25 AGAGAGTAGAAACATTAGAAAACACTACCAATCAGCAAGCAGAAATTAACAGCTATAAAAAATGGCTTTGGAAGACA
 GTGGGCCTAATGTGAACATAGTAACAGACTCTCAATATGCAATGGGAATTTTGACAGCACACCCACACAAAGTG
 ATTCACCATTAGTAGAGCAATTTATAGCCTTAATGATACAAAAGCAACAAATATATTTGCAGTGGGTACCAGCAC
 ATAAAGGAATAGGAGGAAATGAGGAGATAGATAAAATAGTGAGTAAAGGCATTAGAAGAGTTTATTCTTAGAAA
 AAATAGAAGAAGCTCAAGAGAGCATGAAAGATATCATAAATTTGAAAAACCTAGCAGATACATATGGGCTTC
 30 CACAAATAGTAGCAAAAGAGATAGTGGCCATGTGTCCAAAATGTGAGATAAAGGGAGAACCAGTGCATGGACAAG
 TGGATGCCTCACCTGGAACATGGCAGATGGATTGTACTCATCTAGAAGGAAAAGTAGTCATAGTTGCGGTCCATG
 TAGCCAGTGGATTATAGAAGCAGAAGTCATACCTAGGGAAACAGGAAAAAGAAACGGCAAAGTTTCTATTAAAAA
 TACTGAGTAGATGGCCTATAACACAGTTACACACAGACAATGGGCCTAACTTTACCTCCCAAGAAGTGGCAGCAA
 TATGTTGGTGGGGAATAATGAACATACACAGGTATACCATATAACCCCAATCTCAAGGATCAATAGAAAGCA
 35 TGAACAAACAATTAAGAGAGATAATTGGGAAAATAAGAGATGATTGCCAATATACAGAGACAGCAGTACTGATGG
 CTTGCCATATTCACAATTTTAAAGAAAGGGAGGAATAGGGGGACAGACTTCAGCAGAGAGACTAATTAATATAA
 TAACAACACAATTAGAAAATACAACATTTACAAACCAAAATTCAAAAATTTTAAATTTTAGAGTCTACTACAGAG
 AAGGGAGAGACCCTGTGTGGAAGGACCAGCACAATTAATCTGGAAAGGGGAAGGAGCAGTGGTCTCAAGGACG
 GAAGTGACCTAAAGGTTGTACCAAGAAGGAAAGCTAAATTTAAGGATTATGAACCCAAACAAAGAGTGGGTA
 40 ATGAGGGTGACGTGGAAGGTACCAGGGGATCTGATACTAA

SEQ ID NO: 3 Plasmid as defined in Figure 2A (pDNA1 pGM326)

Length: 10528; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..10528; mol_type, other DNA; note, pGM326; organism, synthetic construct
 45 GGTACCTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAAATCAATATTGGCTATTGGCCATT
 GCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTG
 ATTATTGACTAGTTATTAATAGTAATCAATTACGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTAC
 ATAACCTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGT
 50 TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCCTTGGC
 AGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATT
 TGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT
 GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCAT

52

CATCAACAGCATCAGGAAGTTCAGCATTGTGCAGAAAAACCCCTGCAGATGAATGGCATTGAGGAAGATTCTGA
 TGAGCCCCCTGGAGAGGAGACTGAGCCTGGTGCCTGATTCTGAGCAGGGAGAGGCCATCCTGCCTAGGATCTCTGT
 GATCAGCACAGGCCCTACACTGCAGGCCAGAAGGAGGCAGTCTGTGCTGAACCTGATGACCCACTCTGTGAACCA
 GGGCCAGAACATCCACAGGAAAACACAGCCTCCACCAGGAAAGTGAGCCTGGCCCCCTCAGGCCAATCTGACAGA
 5 GCTGGACATCTACAGCAGGAGGCTGTCTCAGGAGACAGGCCTGGAGATTTCTGAGGAGATCAATGAGGAGGACCT
 GAAAGAGTGCTTCTTTGATGACATGGAGAGCATCCCTGCTGTGACCACCTGGAACACCTACCTGAGATACATCAC
 AGTGACACAAGAGCCTGATCTTTGTGCTGATCTGGTGCCTGGTATCTTCTGGCTGAAGTGGCTGCCTCTCTGGT
 GGTGCTGTGGCTGCTGGGAAACACCCCACTGCAGGACAAGGGCAACAGCACCACAGCAGGAACAACAGCTATGC
 TGTGATCATCACCTCCACCTCCAGCTACTATGTGTTCTACATCTATGTGGGAGTGGCTGATACCCTGCTGGCTAT
 10 GGGCTTCTTTAGAGGCCCTGCCCTGGTGCACACACTGATCACAGTGAGCAAGATCCTCCACCACAAGATGCTGCA
 CTCTGTGCTGCAGGCTCCTATGAGCACCCTGAATACCCTGAAGGCTGGGGGCATCCTGAACAGATTCTCCAAGGA
 TATTGCCATCCTGGATGACCTGCTGCTCACCATCTTTGACTTCATCCAGCTGCTGCTGATTGTGATTGGGC
 CATTGCTGTGGTGGCAGTGTGCAGCCCTACATCTTTGTGGCCACAGTGCCTGTGATTGTGGCCTTCATCATGCT
 GAGGGCCTACTTTCTGCAGACCTCCCAGCAGCTGAAGCAGCTGGAGTCTGAGGGCAGAAGCCCCATCTTCACCCA
 15 CCTGGTGACAGCCTGAAGGGCCTGTGGACCCTGAGAGCCTTTGGCAGGCAGCCCTACTTTGAGACCCTGTTCCA
 CAAGGCCCTGAACCTGCACACAGCCAACTGGTTCCTCTACCTGTCCACCCTGAGATGGTTCAGATGAGAATTGA
 GATGATCTTTGTGATCTTTCTCATTGCTGTGACCTTCATCAGCATTCTGACCACAGGAGAGGGAGAGGGCAGAGT
 GGGCATTATCCTGAGCCTGGCCATGAACATCATGAGCACACTGCAGTGGGCAGTGAACAGCAGCATTGATGTGGA
 CAGCCTGATGAGGAGTGTGAGCAGAGTGTCAAGTTCATTGATATGCCACAGAGGGCAAGCCTACCAAGAGCAC
 20 CAAGCCCTACAAGAATGGCCAGCTGAGCAAAGTGATGATCATTGAGAACAGCCATGTGAAGAAGGATGATATCTG
 GCCCAGTGGAGGCCAGATGACAGTGAAGGACCTGACAGCCAAGTACACAGAGGGGGGCAATGCTATCCTGGAGAA
 CATCTCCTTCAGCATCTCCCTGGCCAGAGAGTGGGACTGCTGGGAAGAACAGGCTCTGGCAAGTCTACCCTGCT
 GTCTGCCTTCTGAGGCTGTGAACACAGAGGGAGAGATCCAGATTGATGGAGTGTCTGGGACAGCATCACACT
 GCAGCAGTGGAGGAAGGCCCTTTGGTGTGATCCCCAGAAAGTGTTCATCTTCAGTGGCACCTTCAGGAAGAACC
 25 GGACCCTATGAGCAGTGGTCTGACCAGGAGATTTGAAAAGTGGCTGATGAAGTGGGCTGAGAAGTGTGATTGA
 GCAGTTCCTGGCAAGCTGGACTTTGTCTGGTGGATGGGGGCTGTGTGCTGAGCCATGGCCACAAGCAGCTGAT
 GTGCTGGCCAGATCAGTGTGAGCAAGGCCAAGATCCTGCTGCTGGATGAGCCTTCTGCCACCTGGATCCTGT
 GACCTACCAGATCATCAGGAGGACCTCAAGCAGGCCTTTGCTGACTGCACAGTCATCCTGTGTGACACAGGAT
 TGAGGCCATGTGGAGTGGCAGCAGTTCCTGGTATTGAGGAGAAACAAAGTGAGGCAGTATGACAGCATCCAGAA
 30 GCTGCTGAATGAGAGGAGCCTGTTACAGCAGGCCATCAGCCCCCTCTGATAGAGTGAAGCTGTTCCCCCACAGGAA
 CAGCTCCAAGTGCAAGAGCAAGCCCCAGATTGCTGCCCTGAAGGAGGAGACAGAGGAGGAAGTGCAGGACACCAG
 GCTGTGAGGGCCCAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCC
 TTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTTCAATTTCTC
 CTCCTTGTATAAATCCTGGTGTCTCTTTATGAGGAGTTGTGGCCGTTGTGAGGCAACGTGGCGTGGTGTG
 35 CACTGTGTTTGTGAGCGCAACCCCACTGGTTGGGGCATTGCCACCACCTGTGAGCTCCTTTCCGGGACTTCGCG
 TTTCCCCCTCCCTATTGCCACGGCGGAACCTCATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCTCGGCTGTT
 GGGCACTGACAATTCCGTGGTGTGTGCGGGAAATCATCGTCTTTCCCTTGGCTGCTCGCTGTGTTGCCACCTG
 GATTCTGCGCGGGACGTCTCTGCTACGTCCCTTCGCGCCTCAATCCAGCGGACCTTCCCTCCGCGGGCTGCT
 GCCGGCTCTGCGGCTCTTCCGCGCTTTCGCTTCGCCCTCAGACGAGTCCGATCTCCCTTTGGGCGGCCCTCCCC
 40 GCAAGCTTCGCACTTTTTTAAAGAAAAGGAGGAGTGGATGGGATTTATTACTCCGATAGGACGCTGGCTTGTAA
 CTCAGTCTCTTACTAGGAGACCAGCTTGAGCCTGGGTGTTTCGCTGGTTAGCCTAACCTGGTTGGCCACCAGGGGT
 AAGGACTCCTTGGCTTAGAAAGCTAATAAACTTGCTGCAATTAGAGCTCTTACGCGTCCCGGGCTCGAGATCCGC
 ATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCATCCCGCCCCCTAACTCCGCCAGTTCGCC
 ATCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGACAGAGGCCGAGGCCGCTCGGCCCTGAGCTATTCC
 45 AGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAGCTTATAATGG
 TTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCTAGTCTAGTTGTGGTTGTG
 CAAACTCATCAATGTATCTTATCATGTCTGTCCGCTTCCCTCGCTCACTGACTGCTGCGCTCGCTCGGCTCGGCTG
 CGGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
 ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGC
 50 CCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCCAG
 GCGTTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCGCTTACCGGATACCTGTCCGCTTT
 CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAGCTGAGTGTGAGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCC
 AAGCTGGGCTGTGTGCACGAACCCCCCTTCAGCCGACCGCTGCGCCTTATCCGGTAACATCGTCTTGAGTCC
 AACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
 55 GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTG
 CTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAACAAACACCGCTGGTAGCGGTGGT
 TTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGG
 TCTGACGCTCAGTGGAAACGAAAACTCAGTTAAGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAG
 ATCCTTTTAAATTAATAAAGTGTAAATCAATCAAAAGTATATATGAGTAACTTGGTCTGACAGTTAGAAA
 60 AACTCATCGAGCATCAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGT
 TTCTGTAATGAAGGAGAAAACCTACCGAGGCAGTTCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCG

ACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAAATAAGGTTATCAAGTGAGAAATCACCATGA
 GTGACGACTGAATCCGGTGAGAATGGCAACAGCTTATGCATTTCTTCCAGACTTGTTCAACAGGCCAGCCATTA
 CGCTCGTCATCAAAATCACTCGCATCAACCAACCGTTATTCAATTCGTGATTGCGCCTGAGCGAGACGAAATACG
 CGATCGCTGTTAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACA
 5 ATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTTCCGGGGATCGCAGTGGTGAGTAAC
 CATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAAATCCGTCAGCCAGTTTAGTCTG
 ACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTC
 CCATACAATCGATAGATTGTCGCACCTGATTGCCCCACATTATCGCGAGCCCATTATACCCATATAAATCAGCA
 TCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCCTTGAATATGGCTCATAACACCCCTTGTATTA
 10 CTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGATGATATATTTTATCTTGTGCAATGTACATCAGAGATT
 TGAGACACAACAATTGGTCGACGGATCC

SEQ ID NO: 4 Plasmid as defined in Figure 2B (pDNA1 pGM830)

15 Length: 10536; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..10536; mol_type, other DNA; note, pGM830; organism, synthetic construct
 GGTACCTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATT
 GCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTG
 ATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGCTTAC
 20 ATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATGACGTCAATAATGACGTATGT
 TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC
 AGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTTGGCATT
 TGCCCGAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT
 GATGCGGTTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCAT
 25 TGACGTCAATGGGAGTTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTGCGATCGCCC
 GCCCCGTTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGCTGGCTTGTAACT
 CAGTCTCTTACTAGGAGACCAGCTTGAGCCTGGGTGTTTCGCTGGTTAGCCTAACCTGGTTGGCCACCAGGGGTAA
 GGACTCCTTGGCTTAGAAAGCTAATAAACTTGCCTGCATTAGAGCTTATCTGAGTCAAGTGTCTCATTGACGCC
 TCACTCTCTTGAACGGGAATCTTCTTACTGGGTCTCTCTCTGACCCAGGCGAGAGAACTCCAGCAGTGGCGC
 30 CCGAACAGGGACTTGAGTGAGAGTGTAGGCACGTACAGCTGAGAAGGCGTCGGACGCGAAGGAAGCGCGGGTGC
 GACGCGACCAAGAAGGAGACTTGGTGAGTAGGCTTCTCGAGTGCCGGGAAAAAGCTCGAGCCTAGTTAGAGGACT
 AGGAGAGGCGCTAGCCGTAACCTACTCTGGGCAAGTAGGGCAGGCGGTGGGTACGCAATTGGGGCGGCTACCTCA
 GCACTAAATAGGAGACAATTAGACCAATTTGAGAAAAATACGACTTCGCCCCAACGGAAGAAAAAGTACCAAAAT
 AAACATTTAATATTGGGCAGGCAAGGAGATTGGAGCGCTTCGGCCTCCATGAGAGGTTGTTGGAGACAGAGGAGG
 35 GGTGTAAGAAGATCATAGAAGTCTCTACCCCTAGAACCAACAGGATCGGAGGGCTTAAAAAGTCTGTTCAATC
 TTGTGTGCGTGCTATATTGCTTGCACAAGGAACAGAAAGTGAAAGACACAGAGGAAGCAGTAGCAACAGTAAGAC
 AACACTGCCATCTAGTGGAAAAAGAAAAAGTGCAACAGAGACATCTAGTGGACAAAAAGAAAAATGACAAGGGAA
 TAGCAGCGCCACCTGGTGGCAGTCAGAATTTCCAGCGCAACAACAAGGAAATTGCCTGGGTACATGTACCTTGT
 TCACCGCGCACCTTAAATGCGTGGGTAAAAAGCAGTAGAGGAGAAAAAATTTGGAGCAGAAATAGTACCCATGTTT
 40 CAAGCCCTATCGCCTGCAGGCGCTTTGTGCTAGGGTTCTTAGGCTTCTTGGGGCTGCTGGAATGCATTGGGAG
 CAGCGGCGACAGCCCTGACGGTCCAGTCTCAGCATTGCTTGGTGGGATACTGCAGCAGCAGAAGAATCTGCTGG
 CGGCTGTGGAGGCTCAACAGCAGATGTTGAAGCTGACCATTTGGGGTGTTAAAAACCTCAATGCCCGCGTCACAG
 CCCTTGAGAAAGTACCTAGAGGATCAGGCACGACTAACTCCTGGGGGTGCGCATGGAAACAAGTATGTCATACCA
 CAGTGGAGTGGCCCTGGACAAATCGGACTCCGGATTGGCAAAATAGACTTGGTTGGAGTGGGAAAGACAAATAG
 45 CTGATTTGGAAAGCAACATTACGAGACAATTAGTGAAGGCTAGAGAACAAGAGGAAAAAGAAATCTAGATGCCTATC
 AGAAGTTAACTAGTTGGTCAGATTTCTGGTCTTGGTTCGATTTCTCAAAATGGCTTAACATTTTAAAAAAGGGAT
 TTTTAGTAATAGTAGGAATAATAGGGTTAAGATTACTTTACACAGTATATGGATGTATAGTGAAGGTTAGGCAGG
 GATATGTTCTCTATCTCCACAGATCCATATAAAGCGGCAATTTTAAAAAGAAAGGGAGGAATAGGGGACAGACT
 TCAGCAGAGAGACTAATTAATAATAACAACACAATTAGAAATACAACATTTACAAACCAAAATTCAAAAAAT
 50 TTAAATTTTAGAGCCGCGGAGATCTGTTACATAACTTATGGTAAATGGCCTGCCTGGCTGACTGCCCAATGACCC
 CTGCCCAATGATGTCAATAATGATGTATGTTCCCATGTAATGCCAATAGGGACTTTCCATTGATGTCAATGGGTG
 GAGTATTTATGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTATGCCCCCTATTGATGTCA
 ATGATGGTAAATGGCCTGCCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT

ATGTATTAGTCATTGCTATTACCATGGGAATTCAGTACTGGAGAAGAGCATGCTTGAGGGCTGAGTGCCCTCAG
TGGGCAGAGAGCACATGGCCACAGTCCCTGAGAAGTTGGGGGAGGGGTGGGCAATTGAACTGGTGCCTAGAGA
AGGTGGGGCTTGGGTAACTGGGAAAGTGATGTGGTGTACTGGCTCCACCTTTTTCCCCAGGGTGGGGGAGAACC
ATATATAAGTGCAGTAGTCTCTGTGAACATTCAAGCTTCTGCCTTCTCCCTCCTGTGAGTTTGTAGCCACCATG
5 CAGAGAAGCCCTCTGGAGAAGGCCTCTGTGGTGAGCAAGCTGTTCTTCAGCTGGACCAGGCCCATCCTGAGGAAG
GGCTACAGGCAGAGACTGGAGCTGTCTGACATCTACCAGATCCCCCTCTGTGGACTCTGCTGACAACCTGTCTGAG
AAGCTGGAGAGGGAGTGGGATAGAGAGCTGGCCAGCAAGAAGAACCCCAAGCTGATCAATGCCCTGAGGAGATGC
TTCTTCTGGAGATTTCATGTTCTATGGCATCTTCCCTGTACCTGGGGGAAGTGACCAAGGCTGTGCAAGCCTCTGCTG
CTGGGCAGAAATCATTGCCAGCTATGACCCTGACAACAAGGAGGAGAGGAGCATTGCCATCTACCTGGGCATTGGC
10 CTGTGCCTGCTGTTTCATTGTGAGGACCCTGCTGCTGCACCCTGCCATCTTTGGCCTGCACCACATTGGCATGCGAG
ATGAGGATTGCCATGTTTCAGCCTGATCTACAAGAAAACCTGAAGCTGTCCAGCAGAGTGCTGGACAAGATCAGC
ATTGGCCAGCTGGTGAGCCTGCTGAGCAACAACCTGAACAAGTTTGATGAGGGCCTGGCCCTGGCCCACTTTGTG
TGGATTGCCCTCTGCAGGTGGCCCTGCTGATGGGCCTGATTGGGAGCTGCTGCAGGCCTCTGCCTTTTGTGGC
CTGGGCTTCTGATTGTGCTGGCCCTGTTTCAGGCTGGCCTGGGCAGGATGATGATGAAGTACAGGGACCAGAGG
15 GCAGGCAAGATCAGTGAGAGGCTGGTGATCACCTCTGAGATGATTGAGAATCCAGTCTGTGAAGGCCCTACTGT
TGGGAGGAAGCTATGGAGAAGATGATTGAAAACCTGAGGCAGACAGAGCTGAAGCTGACCAGGAAGGCTGCCTAT
GTGAGATACITTCACAGCTCTGCCTTCTTCTCTGCTGCTTCTTGTGGTGTTCCTGTCTGTGCTGCCCTATGCC
CTGATCAAGGGGATCATCTCTGAGAAGATTTTACCACCATCAGCTTCTGCAATTGTGCTGAGGATGGCTGTGACC
AGACAGTTCCCTGGGCTGTGCAGACCTGGTATGACAGCCTGGGGGCCATCAACAAGATCCAGGACTTCTCTGAG
20 AAGCAGGAGTACAAGACCCTGGAGTACAACCTGACCACCACAGAAGTGGTGATGGAGAATGTGACAGCCTTCTGG
GAGGAGGGCTTTGGGGAGCTGTTTGAGAAGGCCAAGCAGAACAAACAACAGAAAGACCAGCAATGGGGATGAC
TCCCTGTTCTTCTCCAACCTTCTCCCTGCTGGGCACACCTGTGCTGAAGGACATCAACTTCAAGATTGAGAGGGG
CAGCTGCTGGCTGTGGCTGGATCTACAGGGGCTGGCAAGACCAGCCTGCTGATGATGATCATGGGGGAGCTGGAG
CCTTCTGAGGGCAAGATCAAGCACTCTGGCAGGATCAGCTTTTGCAGCCAGTTCAGCTGGATCATGCCTGGCACC
25 ATCAAGGAGAATCATCTTTGGAGTGAGCTATGATGAGTACAGATACAGGAGTGTGATCAAGGCCTGCCAGCTG
GAGGAGGACATCAGCAAGTTTGTGAGAAGGACAACATTGTGCTGGGGGAGGGAGGCATTACACTGTCTGGGGG
CAGAGAGCCAGAATCAGCCTGGCCAGGGCTGTGTACAAGGATGCTGACCTGTACCTGCTGGACTCCCCCTTTGGC
TACCTGGATGTGCTGACAGAGAAGGAGATTTTTGAGAGCTGTGTGTGCAAGCTGATGGCCAACAAGACCAGAATC
CTGGTGACCAGCAAGATGGAGCACCTGAAGAAGGCTGACAAGATCCTGATCCTGCATGAGGGGAGCAGCTACTTC
30 TATGGGACCTTCTCTGAGCTGCAGAACCTGCAGCCTGACTTCAGCTCTAAGCTGATGGGCTGTGACAGCTTTGAC
CAGTTCTCTGCTGAGAGGAGGAACAGCATCCTGACAGAGACCCTGCACAGATTACAGCCTGGAGGGAGATGCCCT
GTGAGCTGGACAGAGACCAAGAAGCAGAGCTTCAAGCAGACAGGGGAGTTTGGGGAGAAGAGGAAGAACTCCATC
CTGAACCCCATCAACAGCATCAGGAAGTTCAGCATTGTGCAGAAAACCCCTGCAGATGAATGGCATTGAGGAA
GATTCTGATGAGCCCCCTGGAGAGGAGACTGAGCCTGGTGCCTGATTCTGAGCAGGGAGAGGCCATCCTGCCTAGG
35 ATCTCTGTGATCAGCACAGGCCCTTACTGCAGGCCAGAAGGAGGCAGTCTGTGCTGAACCTGATGACCCACTCT
GTGAACCAGGGCCAGAATCCACAGGAAAACACAGCCTCCACCAGGAAAGTGAGCCTGGCCCTCAGGCCAAT
CTGACAGAGCTGGACATCTACAGCAGGAGGCTGTCTCAGGAGACAGGCCTGGAGATTCTGAGGAGATCAATGAG
GAGGACCTGAAAGAGTGCTTCTTTGATGACATGGAGAGCATCCCTGCTGTGACCACCTGGAACACCTACCTGAGA
TACATCACAGTGCACAAGAGCCTGATCTTTGTGCTGATCTGGTGCCTGGTGATCTTCTGGCTGAAGTGGCTGCC
40 TCTCTGGTGGTGTGCTGTGGCTGCTGGGAAACACCCCACTGCAGGACAAGGGCAACAGCACCCACAGCAGGAACAAC
AGCTATGCTGTGATCATCACCTCCACCTCCAGCTACTATGTGTTCTACATCTATGTGGGAGTGGCTGATACCCTG
CTGGCTATGGGCTTCTTTAGAGGCCTGCCCTGGTGCACACACTGATCACAGTGAGCAAGATCCTCCACCACAAG
ATGCTGCACTCTGTGCTGCAGGCTCCTATGAGCACCTGAATACCCTGAAGGCTGGGGGCATCCTGAACAGATT
TCCAAGGATATTGCCATCCTGGATGACCTGCTGCCTCTCACCATCTTTGACTTCATCCAGCTGCTGCTGATTGTG
45 ATTGGGGCCATTGCTGTGGTGGCAGTGTGTCAGCCCTACATCTTTGTGGCCACAGTGCCTGTGATTGTGGCCTTC
ATCATGCTGAGGGCTTACTTTCTGCAGACCTCCCAGCAGCTGAAGCAGCTGGAGTCTGAGGGCAGAAGCCCCATC
TTCACCCACCTGGTGACAAGCCTGAAGGGCTGTGGACCCTGAGAGCCTTTGGCAGGCAGCCCTACTTTGAGACC
CTGTTCCACAAGGCCCTGAACCTGCACACAGCCAACCTGTTTCTTACCTGTCCACCCTGAGATGGTTCCAGATG
AGAATTGAGATGATCTTTGTCATCTTCTTCATTGCTGTGACCTTCATCAGCATTCTGACCACAGGAGAGGGAGAG
50 GGCAGAGTGGGCATTATCCTGACCCTGGCCATGAACATCATGAGCACACTGCAGTGGGCAGTGAACAGCAGCATT
GATGTGGACAGCCTGATGAGGAGTGTGAGCAGAGTGTCAAGTTCATTGATATGCCCACAGAGGGCAAGCCTACC
AAGAGCACCAAGCCCTACAAGAATGGCCAGCTGAGCAAAGTGATGATCATTGAGAACAGCCATGTGAAGAAGGAT
GATATCTGGCCAGTGGAGGCCAGATGACAGTGAAGGACCTGACAGCCAAGTACACAGAGGGGGGCAATGCTATC

CTGGAGAACATCTCCTTCAGCATCTCCCTGGCCAGAGAGTGGGACTGCTGGGAAGAACAGGCTCTGGCAAGTCT
ACCCGTGCTGTCTGCCTTCCTGAGGCTGCTGAACACAGAGGGAGAGATCCAGATTGATGGAGTGTCTGGGACAGC
ATCACACTGCAGCAGTGGAGGAAGGCCTTTGGTGTGATCCCCAGAAAGTGTTCATCTTCAGTGGCACCTTCAGG
AAGAACCTGGACCCCTATGAGCAGTGGTCTGACCAGGAGATTTGGAAAGTGGCTGATGAAGTGGGCTGAGAAGT
5 GTGATTGAGCAGTTCCCTGGCAAGCTGGACTTTGTCTGGTGGATGGGGCTGTGTGCTGAGCCATGGCCACAAG
CAGCTGATGTGCTGGCCAGATCAGTGTGAGCAAGGCCAAGATCCTGCTGCTGGATGAGCCTTCTGCCCACCTG
GATCCTGTGACCTACCAGATCATCAGGAGGACCCCTCAAGCAGGCCTTTGCTGACTGCACAGTCATCCTGTGTGAG
CACAGGATTGAGGCCATGCTGGAGTGCCAGCAGTTCCTGGTGATTGAGGAGAACAAAGTGAAGCAGTATGACAGC
ATCCAGAAGCTGCTGAATGAGAGGAGCCTGTTCAAGCAGGCCATCAGCCCTCTGATAGAGTGAAGCTGTTCCCT
10 CACAGGAACAGCTCCAAGTGCAAGAGCAAGCCCCAGATTGCTGCCCTGAAGGAGGAGACAGAGGAGGAAGTGCAG
GACACCAGGCTGTGAGGGCCCAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTAT
GTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTC
ATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTGAGGCAACGTGGC
GTGGTGTGCACTGTGTTTGTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTGAGTCTCTTTCCGGG
15 ACTTTCGCTTTCCCTCCTTATTGCCACGGCGGAATCATCGCCGCTGCTTGGCCGCTGCTGGACAGGGGCT
CGGCTGTGGGCACTGACAATTCCTGTGGTGTGTGCGGGAAATCATCGTCTTTCTTGGCTGCTCGCCTGTGTT
GCCACCTGGATTCTGCGGGGACGTCCTTCTGCTACGTCCTTTCGGCCCTCAATCCAGCGACCTTCTTTCCGCG
GGCTGTGCTGGGCTGCTGCGGCTCTTCCGCTCTTCCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT
GGCTCCCCGCAAGCTTCGCACTTTTTTAAAGAAAAGGGAGGACTGGATGGGATTTATTACTCCGATAGGACGCTG
20 GCTTGTAACTCAGTCTCTTACTAGGAGACCAGCTTGAGCCTGGGTGTTGCTGCTTAGCCTAACCTGGTTGGCCA
CCAGGGGTAAGGACTCCTTGGCTTAGAAAGCTAATAAACTTGCCTGCATTAGAGCTCTTACGCGTCCCGGGCTCG
AGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCCTAACCTCCGCCAG
TTCCGCCCATTTCTCGCCCCATGGCTGACTAATTTTTTTTATTTATGAGAGGCCGAGGCCGCTCGGCTCTGA
GCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTAGGCTTTTGCAAAAAGCTAACTTGTATTATGACGCT
25 TATAATGGTTACAATAAAGCAATAGCATCACAATTTACAAATAAAGCATTTTTTTCAGTGCATTCTAGTTGT
GGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCG
TTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGCTTATCCACAGAATCAGGGGATAACGCAG
GAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
30 GATACCAGGCGTTTTCCCTTGGAAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCTGCGGCTTACCGGATACCTGT
CCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAGCCCGACCGCTGCGCTTATCCGGTAACATATCGTC
TTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
ATGTAGGCGGTGTACAGAGTTCTTGAAGTGGTGGCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
35 GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTA
GCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
CTACGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCT
TCACCTAGATCCTTTTAAATTAATAATGAAGTTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACA
GTTAGAAAAACTCATCGAGCATCAAAATGAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTTGAA
40 AAAGCCGTTTTCTGTAATGAAGGAGAAAACCTACCGAGGCGAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTG
CGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAAT
CACCATGAGTGACGACTGAATCCGGTGAGAATGGCAACAGCTTATGCATTTCTTCCAGACTTGTTCACAGGCC
AGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAACCGTTATTCATTCGTGATTGCGCTGAGCGAGAC
GAAATACGCGATCGCTGTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCG
45 CATCAACAATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTTCCGGGATCGCAGTGG
TGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCGCTCAGCCAGT
TTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCAT
CGGGCTTCCCATACAATCGATAGATTGTGCGACCTGATTGCCCGACATTATCGCGAGCCCATTATACCCATATA
AATCAGCATCCATGTTGGAATTTAATCGCGGCTAGAGCAAGACGTTTCCGTTGAATATGGCTCATAACACCCC
50 TTGTATTACTGTTTATGTAGCAGACAGTTTTTATTGTCATGATGATATATTTTATCTTGTGCAATGTAACATC
AGAGATTTTGAGACACAACAATTGGTCGACGGATCC

SEQ ID NO: 5 Plasmid as defined in Figure 2C (pDNA2a pGM691)

Length: 9064; Molecule Type: DNA; Features Location/Qualifiers: source, 1..9064; mol_type, other DNA; note, pGM691; organism, synthetic construct

5 ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCG
TTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGT
ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGC
ATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
10 TGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTTGTATTAT
TTATTTTTTAAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGG
GCGAGGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTT
TTATGGCGAGGCGGGCGGGCGGGCGGCCCTATAAAAAGCGAAGCGCGCGGGCGGGGAGTCGCTGCGCGCTGCC
TTCGCCCCGTGCCCGCTCCGCCGCCGCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCTTACTCCACAGG
15 TGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAAATTAGCGCTTGGTTTAAATGACGGCTTGTTCCTTTCTGT
GGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGAGCGGCTCGGGGGGTGCGTGCGTGT
GTGTGTGCTGGGGAGCGCGCGTGCAGGCTCCGCGCTGCCCGCGGCTGTGAGCGCTGCGGGCGCGCGCGGGGC
TTTGTGCGCTCCGCGAGTGTGCGCGAGGGGAGCGCGCGGGGGCGGTGCCCGCGGTGCGGGGGGGGCTGCGAGG
GGAACAAAGGCTGCGTGCGGGGTGTGTGCTGCGGGGGTGTGAGCAGGGGGTGTGGGCGCGTGTGCGGGCTGCAAC
20 CCCCCCTGCACCCCCCTCCCGAGTTGCTGAGCAGCGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCG
CGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCTCGGGCGGGG
AGGGCTCGGGGAGGGGCGCGGCGGCCCGGAGCGCGGCGGCTGTGAGGCGCGGCGAGCCGAGCCATTGCC
TTTTATGTAATCGTGCGAGAGGGCGCAGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGC
GCCGCCGACCCCCCTCTAGCGGGCGGGGGCGAAGCGGTGCGGCGCGGCGAGGAAGGAAATGGGCGGGAGGGCC
25 TTCGTGCGTGC CGCGCGCCGCTCCCTTCTCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGACGGCTGCCTTC
GGGGGGGACGGGGCAGGGCGGGGTTCCGCTTCTGCGGTGTGACCGCGGCTCTAGAGCCTCTGCTAACCATGTTT
ATGCCTTCTTCTTTTCTTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAAT
TGCTCGAGCCACCATGGGAGCTGCCACATCTGCCCTGAATAGACGGCAGCTGGACAGTTCGAGAAGATCAGACT
GCGGCCCCAACGCCAAGAAGAAGTACCAGATCAAGCACCTGATCTGGGCCGGCAAAGAGATGGAAGATTGCGCCT
30 GCACGAGCGGCTGCTGGAACCGAGGAAGGCTGCAAGAGAATTATCGAGGTGCTGTACCCTCTGGAACCTACCGG
CTCTGAGGGCTGAAGTCCCTGTTCAATCTCGTGTGCGTGTGTACTGCCTGCACAAAGAACGAAAGTGAAGGA
CACCGAAGAGGCCGTGGCCACAGTTAGACAGCACTGCCACCTGGTGGAAAAAGAGAAGTCCGCCACAGAGACAAG
CAGCGGCCAGAAGAAGAACGACAAGGGAATTGCTGCCCTCCTGGCGGCAGCCAGAATTTCTGCTCAGCAGCA
GGGAAACGCTGGGTGCACGTTCCACTGAGCCCTAGAACACTGAATGCCTGGGTCAAAGCCGTGGAAGAGAAGAA
35 GTTTGGCGCCGAGATCGTGCCCATGTTCCAGGCTCTGTCTGAGGGCTGCACCCCTTACGACATCAACCAGATGCT
GAACGTGCTGGGAGATCACCAGGGCGCTCTGCAGATCGTGAAAGAGATCATCAACGAAGAGGCTGCCAGTGGGA
CGTGACACATCCATTGCTGCTGGACCTCTGCCAGCCGACAACTGAGAGATCCTAGAGGCTCTGATATCGCCGG
CACCACCAGCTCTGTGCAAGAGCAGCTGGAATGGATCTACACCGCCAATCCTAGAGTGGACGTGGGCGCCATCTA
CAGAAGATGGATCATCCTGGGCTGCAGAAATGCGTGAAGATGTACAACCCCGTGTCCGTGCTGGACATCAGACA
40 GGGACCCAAAGAGCCCTTCAAGGACTACGTGGACCGTTCTATAAGGCCATTAGAGCCGAGCAGGCCAGCGCGGA
AGTGAAGCAGTGGATGACAGAGAGCCTGCTGATCCAGAACGCCAATCCAGACTGCAAAGTGATCCTGAAAGGCCT
GGGCATGCACCCACACTGGAAGAGATGCTGACAGCCTGTCAAGGCGTTGGCGGCCCTTCTTACAAAGCCAAAGT
GATGGCCGAGATGATGCAGACCATGCAGAACCAGAACATGGTGCAGCAAGGCGGCCCTAAGAGACAGAGGCTCC
TCTGAGATGCTACAACCTCGCGCAAGTTCCGGCCACATGCAGAGACAGTGTCTGAGCCTAGGAAAAACAAATGTCT
45 AAAGTGTGGAATAATGGGACACCTAGCAAAAGACTGCAGGGGACAGGTGAATTTTTTAGGGTATGGACGGTGGAT
GGGGGCAAAACCGAGAAATTTCCCGCGCTACTCTGGAGCGGAACCGAGTGCCTCCTCCACCGAGCGGCAC
CACCCCATACGACCCAGCAAAAGACTCCTGCAGCAATATGCAGAGAAAGGGAACAACTGAGGGAGCAAAAGAG
GAATCCACCGCAATGAATCCGGATTGACCGAGGGATATTCTTTGAACTCCCTCTTTGGAGAAGACCAATAAAG
ACCGTGTACATCGAGGGCGTGGCCATCAAGGCTCTGCTGGATACAGGCGCCGACGACACCATCATCAAGAGAAC
50 GACCTGCAGCTGAGCGGCCCTTGGAGGCTAAGATCATTGGAGGAATCGGCGGAGGCTGAACGTCAAAGAGTAC
AACGACCGGAAGTGAAGATCGAGGACAAGATCCTGAGGGGCACAACTCTGCTGGGCGCCACCTATCAACATC
ATCGGCAGAAATCTGCTGGCCCTGCCGGCGCTAGACTGGTTATGGGACAGCTCTCTGAGAAGATCCCCGTGACA

CCCGTGAAGCTGAAAGAAGGCGCTAGAGGACCTTGTGTGCGACAGTGGCCTCTGAGCAAAGAGAAGATTGAGGCC
CTGCAAGAAATCTGTAGCCAGCTGGAACAAGAGGGCAAGATCAGCAGAGTTGGCGGCGAGAACGCCTACAATACC
CCTATCTTCTGCATCAAGAAAAAGGACAAGAGCCAGTGGCGGATGCTGGTGGACTTTAGAGAGCTGAACAAGGCT
ACCCAGGACTTCTTCGAGGTGCAGCTGGGAATTCCTCATCTGCGGCGCTGCGGAAGATGAGACAGATCACAGTG
5 CTGGATGTGGGCGACGCCTACTACAGCATCCCTCTGGACCCCAACTTCAGAAAGTACACCGCCTTCACAATCCCC
ACCGTGAACAATCAAGGCCCTGGCATCAGATACCAGTTCAACTGCCTGCCTCAAGGCTGGAAGGGCAGCCCCACC
ATTTTTTCAGAAATACCGCGCCAGCATCTTGGGAAGAAATCAAGAGAAACCTGCCTGCTCTGACCATCGTGCAGTAC
ATGGACGATCTGTGGGTTCGGAAGCCAAAGAGAATGAGCACACCCACGACAAGCTGGTGGAAACAGCTGAGAACAAAG
CTGCAGGCCCTGGGGCCTCGAAACCCCTGAGAAGAAGGTGCAGAAAGAACCTCCTTACGAGTGGATGGGCTACAAG
10 CTGTGGCCTCACAAGTGGGAGCTGAGCCGGATTACAGCTCGAAGAGAAGGACGAGTGGACCGTGAACGACATCCAG
AAACTCGTGGGCAAGCTGAATTGGGCAGCCCAGCTGTATCCCGGCGCTGAGGACCAAGAATCTGCAAGCTGATC
CGGGGAAAGAAGAACCTGCTGGAACCTGGTCACATGGACACCTGAGGCCGAGGCCGAATATGCCGAGAATGCCGAA
ATCCTGAAAACCGAGCAAGAGGGGACCTACTACAAGCCTGGCATTCCAATCAGAGCTGCCGTGCAGAACTGGAA
GGCGGCCAGTGGTCTTACCAGTTTAAGCAAGAAGGCCAGGTCTGAAAGTGGGCAAGTACACCAAGCAGAAGAAC
15 ACCCACACCAACGAGCTGAGGACACTGGCTGGCCTGGTCCAGAAAATCTGCAAGAGGCCCTGGTCATTTGGGGC
ATCCTGCCTGTTCTGGAACCTGCCCATTGAGCGGGAAGTGTGGGAACAGTGGTGGGCCGATTACTGGCAAGTGTCT
TGGATCCCCGAGTGGGACTTCGTGTCTACCCCTCCTCTGCTGAAACTGTGGTACACCCCTGACAAAAGAGCCCAT
CCTAAAGAGGACGTCTACTACGTTGACGGCGCCTGCAACCGGAATCCAAAGAAGGCAAGGCCGGCTACATCAGC
CAGTACGGCAAGCAGAGAGTGGAAACCTGGAACCAACCAACCAAGCAGGCGGAGCTGACCGCCATTAAAGATG
20 GCCCTGGAAGATAGCGGCCCAATGTGAACATCGTGACCGACTCTCAGTACGCCATGGGAATCCTGACAGCCAG
CCTACACAGAGCGATAGCCCTCTGGTTGAGCAGATCATTGCCCTGATGATTAGAAGCAGCAAATCTACCTGCAG
TGGGTGCCCGCTCACAAAGGCATCGGCGGAAACGAAGAGATCGATAAGCTGGTGTCCAAGGGAATCAGACGGGTG
CTGTTCTTGAAAAAGATTGAAGAGGCCCAAGAGGAACACGAGCGCTACCACAACAACCTGGAAGAATCTGGCCGAC
ACCTACGGACTGCCCCAGATCGTGGCCAAAGAAATCGTGGCTATGTGCCCCAAGTGTGAGATCAAGGGCGAACCT
25 GTGCACGGCAAGTGGATGCTTCTCCTGGCACATGGCAGATGGACTGTACCCACCTGGAAGGCAAGTGGTCATC
GTGGCTGTGCACGTGGCTCCGGCTTTATTGAGGCCGAAGTGATCCCCAGAGAGACAGGCAAGAAACCGCCAAAG
TTCTGTCTGAAGATCCTGTCCAGATGGCCCATCACACAGCTGCACACCGACAACGGCCCTAACTTCACATCTCAA
GAGGTGGCCGCCATCTGTTGGTGGGAAAGATTGAGCACACAACCGGCATTCCCTACAATCCACAGAGCCAGGGC
AGCATCGAGTCCATGAACAGCAGCTCAAAGAGATTATCGGCAAGATCCGGGACGACTGCCAGTACACAGAAACA
30 GCCGTGCTGATGGCCTGTACATCCACAACCTTCAAGCGGAAAGGCGGCATCGGAGGACAGACATCTGCCAGAGA
CTGATCAATATCATCACCCTCAGCTGGAAATCCAGCACCTCCAGACCAAGATCCAGAAGATTCTGAACCTCCGG
GTGTACTACCGCGAGGGCAGAGATCCTGTTTGAAAGGCCCAGCACAGCTGATCTGGAAAGGCGAAGGTGCCGTG
GTGCTGAAGGATGGCTCTGATCTGAAGGTGGTGGCCAGACGGAAGGCCAAGATTATCAAGGATTACGAGCCCAAA
CAGCGCGTGGGCAATGAAGGCGACGTTGAGGGCACAAGAGGCAGCGACAATTGAAATTCACCTCCTCAGGTGCAGG
35 CTGCCTATCAGAAGGTGGTGGCTGGTGTGGCCAAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCCCTCTG
CCAAAAATTATGGGGACATCATGAAGCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTG
CAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAATCATTTAAACATCAGA
ATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAG
GTCATCAGTATATGAACAGCCCCCTGTGTCCATTCTTATTCATAGAAAAGCCTTGACTTGAGGTTAGATTT
40 TTTTATATTTTGTGTTTGTGTTATTTTTTTCTTTAACATCCCTAAAATTTTCTTACATGTTTTACTAGCCAGAT
TTTTCTCTCTCCTGACTACTCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCA
AGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATACGA
GCCGGAAGCATAAAGTGTAAGCCTGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTG
CCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGCGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCC
45 TAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCGCCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTA
TTTATGCAGAGGCCGAGGCCGCTCGGCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGTCAAAAAGCTAACTTGTATTGTCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTTAC
AAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTCC
GCTTCTCTGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTA
50 ATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAAC
CGTAAAAAGGCCGCGTGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCA
AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCT
CCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC

TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCTGCTCCAGCTGGGCTGTGTGCACGAACCCCCCGTTCAG
 CCCGACCGCTGCGCCTTATCCGGTAACATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA
 GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAAC
 TACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGT
 5 AGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGA
 AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAA
 GGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCA
 ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTA
 TTCATATCAGGATTATCAATACCATATTTTGAAGAGCCGTTTCTGTAATGAAGGAGAAAACCTCACCGAGGCGAG
 10 TTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTT
 CCTCGTCAAAAAATAAGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAACAGC
 TTATGCATTTCTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAA
 CCGTTATTTCATTTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGA
 ATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTCTAAT
 15 ACCTGGAATGCTGTTTTTCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTG
 ATGGTCGGAAGAGGCATAAATTCGCTCAGCCAGTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTA
 CCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACTGATTGC
 CCGACATATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGCCCTAGAGCAA
 GACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCAT
 20 GATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACAATTGGTCGAC

SEQ ID NO: 6 Plasmid as defined in Figure 2D (pDNA2b pGM299)

Length: 3384; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..3384; mol_type, other DNA; note, pGM299; organism, synthetic construct
 25 TCAATATTGGCCATTAGCCATATTATTTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATAC
 GTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGGCATTGATTATT
 GACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACT
 TACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCAT
 30 AGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACA
 TCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCA
 GTACATGACCTTACGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG
 GTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGT
 CAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCGCCCCGTTGACGCA
 35 AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCACTAGAAG
 CTTTATTGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGCTTCTGACACAACAGTCTCGAATTAAG
 CTGCAGAAGTTGGTCTGAGGCATGGGCAGGTAAATCAAGGTTACAAGACAGGTTTAAAGGACCAATAGAA
 ACTGGGCTTGTGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTGC
 CTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAAGGCTAGAGTACTTAATACGACTCACTATA
 40 GGCTAGCCTCGAGAATTCGATTATGCCCTAGGACCAAGAAAGAAAGATTGCTTCGCTTGATTGGCTCCTTTA
 CAGCACCAATCCATATCCACCAAGTGGGGAAGGGACGGCCAGACAACGCCGACGAGCCAGGAGAAGGTGGAGACA
 ACAGCAGGATCAAATTAGAGTCTTGGTAGAAAGACTCCAAGAGCAGGTGTATGCAGTTGACCGCCTGGCTGACGA
 GGCTCAACACTTGGCTATACAACAGTTGCCTGACCCCTCTCATTACGCTTAGAATCACTAGTGAATTCACGCGTG
 GTACCTCTAGAGTCGACCCGGGCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTGGACAAACCAC
 45 AACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTTATAAG
 CTGCAATAAACAAGTTAACAACAACAATTGCATTCTTTATGTTTCAGGTTTCAGGGGAGATGTGGGAGGTTTT
 TTAAAGCAAGTAAACCTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGACCAATTGTTGTGTCTCAAAATC
 TCTGATGTTACATTGCACAAGATAAAAAATATATCATCATGAACAATAAACTGTCTGCTTACATAAACAGTAATA
 CAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTTCTTGCTCTAGGCCGCGATTAAATTCACATGGATGCTG
 50 ATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGC
 CCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGAC
 TAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTAC
 TCACCACTGCGATCCCCGAAAAACAGCATTCCAGGTATTAGAAGAAATATCCTGATTACAGTGAAAAATATTGTTG
 ATGCGCTGGCAGTGTCTTCTGCGCCGGTTGCATTTCGATTCCTGTTGTAATTGTCCTTTAACAGCGATCGCGTAT
 55 TTCGTCTCGCTCAGGCGCAATCAGAAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATG
 GCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTGTTGCCATTCTCACCGGATTACGTCGTCATCATG

GTGATTTCTCACTTGATAACCTTATTTTGGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAA
TCGCAGACCGATAACCAGGATCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGC
TTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTGATGCTCGATGAGTTTTCT
5 AACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTTAATTTAAAAAGGATCTAGG
TGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCG
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAACAC
CGCTACCAGCGGTGGTTTTGTTGCGCGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACGGCTTCAGCAGAG
CGCAGATAACAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCAGCCTA
10 CATACCTCGCTCTGCTAATCCTGTTACCACTGGCTGCTGCCAGTGGCGATAAGTCTGTCTTACCGGGTTGGACT
CAAGACGATAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACAGCCAGCTTGGAGC
GAACGACCTACACCGAAGTGAATACCTACAGCGTGAAGTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGG
CGGACAGGTATCCCGTAAGCGCGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACCGCTGGT
ATCTTTATAGTCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
15 GCCTATGAAAAACGCCAGCAACGCGGCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGGCTC
GACAGATCT

SEQ ID NO: 7 Plasmid as defined in Figure 2E (pDNA3a pGM301)

Length: 6264; Molecule Type: DNA; Features Location/Qualifiers: source,
20 1..6264; mol_type, other DNA; note, pGM301; organism, synthetic construct
ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCG
TTACATAACTTACGGTAAATGGCCCGCTGGCTGACGCCCCAACGACCCCGCCATTGACGTCAATAATGACGT
ATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
25 TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGC
ATTATGCCCAGTACATGACCTTATGGGACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
TGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCCAATTTTGTATTAT
TTATTTTTTAAATATTTTGTGACGCGATGGGGGCGGGGGGGGGGGGGGGGGGGCGCGCCAGGCGGGGCGGGGCGGG
GCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTT
30 TTATGGCGAGGCGGGCGGGCGGGCGGGCCCTATAAAAAGCGAAGCGCGGGCGGGGAGTCTGCTGCGCGCTGCC
TTCGCCCCGTGCCCGCTCCGCGCGCGCTCGCGCGCCCCCGCGCTCTGACTGACCGCTTACTCCACAGG
TGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAAAGACGGCTTGTCTTTTCTGT
GGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGAGCGGCTCGGGGGTGCCTGCGTGT
GTGTGTGCGTGGGGAGCGCGCGTCCGCGCTGCCCGCTGCCCGGGCGCTGTGAGCGCTGCGGGGCGGGCGGGGCG
35 TTTGTGCGCTCCGCGAGTGTGCGCGAGGGGAGCGCGGCGGGGGCGGTGCCCGCGGTGCGGGGGGGGCTGCGAGG
GGAACAAAGGCTGCGTGCAGGGTGTGTGCGTGGGGGGGTGAGCAGGGGTGTGGGCGCGTCCGTGCGGGCTGCAAC
CCCCCTGACCCCCCTCCCGAGTTGCTGAGCAGGCGCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCG
CGGGGCTCGCCGTGCGGGGCGGGGGGTGGCGGCAGGTGGGGGTGCGGGGCGGGGCGGGGCGCCCTCGGGCCGGG
AGGGCTCGGGGAGGGGCGCGGGCGGGCCCCGAGCGCGGCGGCTGTGAGGCGCGGCGAGCCGAGCCATGCGC
40 TTTTATGGTAATCGTGCAGAGGGGCGCAGGACTTCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGC
GCCGCGCACCCCCCTCTAGCGGGCGGGGGCGAAGCGGTGCGGCGCGGCGAGGAAGGAAATGGGCGGGGAGGGCC
TTCGTGCGTGCAGCGCGCGCGCTCCCTTCTCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGACGGCTGCCTTC
GGGGGGGACGGGGCAGGGCGGGGTTTCGGCTTCTGGCGTGTGACCGCGGCTCTAGAGCCTCTGCTAACCATGTT
ATGCTTCTTCTTTTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAAT
TCGATTGCCATGGCAACATATATCCAGAGAGTACAGTGCATCTCAACATCACTACTGGTTGTCTCACCACATTG
45 GTCTCGTGTGAGATTCCAGGGATAGGCTCTCTAACATAGGGGTCTAGTTCGATGAAGGGAATCACTGAAGATA
GCTGGATCCCAGCAATCGAGGTACATAGTACTGAGTCTAGTTCCGGGGGTAGACTTTGAGAATGGGTGCGGAACA
GCCCAGGTTATCCAGTACAAGAGCCTACTGAACAGGCTGTTAATCCCATTGAGGGATGCCTTAGATCTTCAGGAG
GCTCTGATAACTGTCACCAATGATACGACACAAAATGCCGGTGTCCCCAGTCGAGATTCTTCGGTGTGTGATT
GGTACTATCGCACTTGGAGTGGCGACATCAGCACAAATCACCGCAGGGATTGCACTAGCCGAAGCGAGGGAGGCC
50 AAAAGAGACATAGCGCTCATCAAAGAATCGATGACAAAAACACACAAGTCTATAGAAGTGTGCAAAACGCTGTG
GGGGAACAAATTTGCTCTAAAGACTCCAGGATTTTCGTGAATGATGAGATCAAACCGCAATAAGCGAATTA
GGCTGTGAGACTGCTGCCCTAAGACTGGGTATAAAATTGACACAGCATTACTCCGAGCTGTTAACTGCGTTCCGC
TCGAATTTTCGAACCATCGAGAGAAGAGCCTCACGCTGCAGGCGCTGTCTTCACTTTACTCTGCTAACATTACT
GAGATTATGACCACAATCAGGACAGGGCAGTCTAACATCTATGATGTCATTTATACAGAACAGATCAAAGGAACG

GTGATAGATGTGGATCTAGAGAGATACATGGTCACCCTGTCTGTGAAGATCCCTATTCTTTCTGAAGTCCCAGGT
GTGCTCATACACAAGGCATCATCTATTCTTACAACATAGACGGGAGGAATGGTATGTGACTGTCCCCAGCCAT
ATACTCAGTCGTGCTTCTTTCTTAGGGGGTGCAGACATAACCGATTGTGTTGAGTCCAGATTGACCTATATATGC
CCCAGGGATCCCGCACAACCTGATACCTGACAGCCAGCAAAAGTGTATCCTGGGGGACACAACAAGGTGTCTGTG
5 ACAAAAGTTGTGGACAGCCTTATCCCCAAGTTTGCTTTTGTGAATGGGGCGTTGTTGCTAACTGCATAGCATCC
ACATGTACCTGCGGGACAGGCCGAAGACCAATCAGTCAGGATCGCTCTAAAGGTGTAGTATTCTTAACCCATGAC
AACTGTGGTCTTATAGGTGTCAATGGGGTAGAATTGTATGCTAACCGGAGAGGGCACGATGCCACTTGGGGGGT
CAGAAGTTGACAGTCGGTCTGCAATTGCTATCAGACCCGTTGATATTTCTCTCAACCTTGCTGATGCTACGAAT
TTCTTGCAAGACTCTAAGGTGAGCTTGAGAAAGCAGGAAAATCCTCTCGGAGGTAGGTAGATGGTACAACCTCA
10 AGAGAGACTGTGATTACGATCATAGTAGTTATGGTCGTAATATTGGTGGTCATTATAGTGATCATCATCGTGCTT
TATAGACTCAGAAGGTGAAATCACTAGTGAATTCACCTCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGG
TGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCATGAA
GCCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTG
TCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTAAAAACATCAGAATGAGTATTGGTTTAGAGTTTGGC
15 AACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCC
TGCTGTCCATTCTTATTCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTGTTGTTATT
TTTTTCTTTAACATCCCTAAAATTTTCTTACATGTTTTTACTAGCCAGATTTTTCTCTCTCCTGACTACTCCC
AGTCATCTGTCCTCTTCTCTTATGGAGATCCCTGCACCTGCAGCCCAAGCTTGGCCTAATCATGGTCAATAGC
TGTTTCTGTGTGAAATTTGTATCCGCTCACAAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCT
20 GGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
CGTGCCAGCGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCCGCCCTAAC
TCCGCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTC
GGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCTAGGCTTTTGCAAAAAGCTAACTTGTTT
ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCAT
25 TCTAGTTGTGGTTGTCCAACTCATCAATGTATCTTATCATGTCTGTCCGCTTCTCGCTCACTGACTCGCTGC
GCTCGGTGCTTCCGCTGCGGCGAGCGGTATCAGCTCAAGGCGGTAATACGGTTATCCACAGAATCAGGGG
ATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAGGCCGCGTTGCTGGCGT
TTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCTGCCGCTTACCG
30 GATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGG
TGAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCTTCAGCCCGACCGCTGCCCTTATCCGGTA
ACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCA
GAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTAT
TTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCA
35 CCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACGGGTCTGACGCTCAGTGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATAGATAAACTT
GGTCTGACAGTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTATATCAGGATTATCAATACCAT
ATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGT
40 ATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCTCGTCAAAAAATAAGGTTATCAA
GTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAACAGCTTATGCATTTCTTCCAGACTTGTT
CAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAACCGTTATTCTTCGTGATTGCGCCT
GAGCGAGACGAAATACGCGATCGCTGTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACA
CTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTTCCGGGGA
45 TCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTGGGAAGAGGCATAAATTCCG
TCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACT
CTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCCACATTATCGCGAGCCCATTTAT
ACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCA
TAACACCCCTTGATTACTGTTTATGTAAGCAGACAGTTTTATTGTTTCATGATGATATATTTTTATCTTGTGCAA
50 TGTAACATCAGAGATTTTGAGACACAACAAATTGGTCGAC

SEQ ID NO: 8 Plasmid as defined in Figure 2F (pDNA3b pGM303)

Length: 6522; Molecule Type: DNA; Features Location/Qualifiers: source, 1..6522; mol_type, other DNA; note, pGM303; organism, synthetic construct

5 ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGG
TTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGT
ATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGC
ATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
10 TGGTCGAGGTGAGCCCCACGTTCTGTCTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTTGTATTAT
TTATTTTTTAAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGG
GCGAGGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTT
TTATGGCGAGGCGGGCGGGCGGGCGGCCCTATAAAAAGCGAAGCGCGCGGGCGGGAGTCTGCTGCGCGCTGCC
TTCGCCCCGTGCCCGCTCCGCGCGCCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCTTACTCCACAGG
TGAGCGGGCGGGACGGCCCTTCTCTCCGGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTCCTTTCTGT
15 GGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGAGCGGCTCGGGGGGTGCGTGCGTGT
GTGTGTGCGTGGGGAGCGCGCGTGCAGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGCGCGGGG
TTTGTGCGCTCCGCGAGTGTGCGGAGGGGAGCGCGCGGGGGGGGGTGTGCCCCGCGTGTGCGGGGGGGCTGCGAGG
GGAACAAAGGCTGCGTGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGTGTGCGGCTGCAAC
CCCCCTGACCCCCCTCCCGAGTTGCTGAGCACGCCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCG
20 CGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCGGGGCGGGGCGGGGCGCCCTCGGGCCGGGG
AGGGCTCGGGGGAGGGGCGCGGCGGCCCGGAGCGCGGCGGCTGTGAGGCGCGGCGAGCCGAGCCATTGCC
TTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGC
GCCGCCGACCCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCGGCGAGGAAGGAAATGGGCGGGGAGGGCC
TTCGTGCGTGTGCGCGCGCGCGCTCCCTTCTCCCTCTCCAGCTTCGGGGCTGTCCGCGGGGGGACGGGGCAGGGC
25 GGGGTTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCATGTTTCATGCTTCTTCTTTTCTCT
ACAGCTCCTGGGCAACGTGCTGGTTATGTGCTGTCTCATCTTTTGGCAAAGAATCTCTCGAGCATGTGGTCTG
AGTTAAAAATCAGGAGCAACGACGAGGTGAAGGACCAGAGGACGCCAACGACCCCGGGGAAAGGGGGTGAAC
ACATCCATATCCAGCCATCTCTACCTGTTTATGGACAGAGGGTTAGGGATGGTGATAGGGGCAACGTCATCGT
ACTGGTCTACTTCTCTAGTGGTAGCACCACAAAACAGCATCAGGTTGGGAGAGGTCAAGTAAAGCCGACAT
30 GGTTGCTGATTCTCTCATTCACCCAGTGGGCTTTGTCAATTGCCACAGTGATCATCTGTATCATATTTCTGCTA
GACAAGGGTATAGTATGAAAGAGTACTCAATGACTGTAGAGGCATTGAACATGAGCAGCAGGGAGGTGAAAGAGT
CACTTACCAGTCTAATAAGGCAAGAGGTTATAGCAAGGGCTGTCAACATTGAGGCTCTGTGCAAACCGGAATCC
CAGTCTTGTGAACAAAAACAGCAGGGATGTCTCCAGATGATTGATAAGTCTGTCAGCAGACAAGAGCTCACTC
AGCACTGTGAGAGTACGATCGCAGTCCACCATGCCGATGGAATTGCCCCACTTGAGCCACATAGTTTCTGGAGAT
35 GCCCTGTGCGGAGAACCGTATCTTAGCTCAGATCCTGAAATCTCATTGCTGCCTGGTCCGAGCTTGTATCTGGTT
CTACAACGATCTCTGGATGTGTTAGGCTCCCTTCACTCTCAATTGGCGAGGCAATCTATGCCTATTCATCAAAATC
TCATTACACAAGGTTGTGCTGACATAGGGAATCATATCAGGTCTGCGAGCTAGGGTACATATCACTCAATTGAG
ATATGTTCCCTGATCTTAACCCGTTAGTGTCCCACACTTATGACATCAACGACAATCGGAAATCATGCTCTGTGG
TGGCAACCGGACTAGGGGTTATCAGCTTTGCTCCATGCCGACTGTAGACGAAAGAACCGACTACTCTAGTGATG
40 GTATTGAGGATCTGGTCCCTGATGTCTGGATCTCAAAGGGAGAACTAAGTCTCACCAGGATCGCAACAGCGAGG
TAGATCTTGATCACCCGTTCTCTGCACTATACCCAGTGTAGGCAACGGCATTGCAACAGAAGGCTCATTGATAT
TTCTTGGGTATGGTGGACTAACCACCCCTCTGCAGGGTGATACAAAATGTAGGACCCAAGGATGCCAACAGGTGT
CGCAAGACACATGCAATGAGGCTCTGAAAATTACATGGCTAGGAGGGAAACAGGTGGTCTAGCGTGATCATCCAGG
TCAATGACTATCTCTCAGAGAGGCCAAAGATAAGAGTCACAACCATTCCAATCACTCAAACTATCTCGGGGCGG
45 AAGGTAGATTATTAATAATGGGTGATCGGGTGTACATCTATACAAGATCATCAGGCTGGCACTCTCAACTGCAGA
TAGGAGTACTTGATGTGACGCCACCTTTGACTATCAACTGGACACCTCATGAAGCCTTGTCTAGACCAGGAAATA
AAGAGTGCAATTGGTACAATAAGTGTCCGAAGGAATGCATATCAGGCGTATACACTGATGCTTATCCATTGTCCC
CTGATGCAGCTAACGTCGCTACCGTCACGCTATATGCCAATACATCGCGTGTCAACCAACAAATCATGTATTCTA
ACACTACATACATATAAATATGTTAAGGATAAAGGATGTTCAATTAGAGGCTGCATATACCACGACATCGTGTA
50 TCACGCATTTTGGTAAAGGCTACTGTCTTACATCATCGAGATCAATCAGAAGAGCCTGAATACCTTACAGCCGA
TGCTCTTTAAGACTAGCATCCCTAAATTATGCAAGGCCGAGTCTTAAGCGGCCGCGCATGCGAATTCACCTCTCA
GGTGACGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGCAATGCCCTGGCTCACAATACCACTGAGATCTTTT
TCCCTCTGCCAAAAATTATGGGGACATCATGAAGCCCTTGAAGCATCTGACTTCTGGCTAATAAAGGAAATTTAT

TTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAAA
 ACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCT
 ATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGTGTCTATTCTTATTCCATAGAAAAGCCTTGACTTGAGG
 TTAGATTTTTTTTATATTTTGTGTTATTTTTTCTTTAACATCCCTAAAATTTTCTTACATGTTTTACT
 5 AGCCAGATTTTTCTCTCTCTCTGACTACTCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCT
 GCAGCCCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACA
 ACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGC
 GCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCTGTCCAGCGGATCCGCATCTCAATTAGTCAGCAACCATACT
 CCGCCCCCTAACTCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTTCTCCGCCCATGGCTGACTAAT
 10 TTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGG
 AGGCCTAGGCTTTTGCAAAAGCTAACTTGTTTATTGCAGCTTATAATGGTTACAAAATAAGCAATAGCATCACA
 AATTTACAAAATAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCAT
 GTCTGTCCGCTTCTCTCGCTCACTGACTCGCTGCGCTCGGTCTGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
 AGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
 15 CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATC
 GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCTTGGAAGCTCCCTCG
 TCGCTCTCTCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTT
 CTCATGCTCAGCTCAGCTGTAGTATCTCAGTTCCGTTGAGTTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
 CCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACCTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
 20 CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT
 GGCCTAACCTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAAA
 GAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTA
 CGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACT
 CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGTT
 25 TTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTAGAAAACTCATCGAGCATCAAATGAACT
 GCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCAC
 CGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTA
 TTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATG
 GCAACAGCTTATGCATTTCTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCAT
 30 CAACCAAACCGTTATTCACTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTAC
 AAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATT
 CTTCTAATACCTGGAATGCTGTTTTTCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAA
 AATGCTTGATGGTCGGAAGAGGCATAAATCCGTCAGCCAGTTTGTCTGACCATCTCATCTGTAAACATCATTGG
 CAACGCTACCTTTGCCATGTTTCAGAAACAACCTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTGCGCAC
 35 CTGATTGCCGACATTATCGCGAGCCCATTTATACCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCC
 TAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTAATTACTGTTTATGTAAGCAGACAGTTTA
 TTGTTTATGATGATATATTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACAATTTGGTCGAC

40 **SEQ ID NO: 9 Plasmid as defined in Figure 2G (pDNA2a pGM297)**

Length: 9886; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..9886; mol_type, other DNA; note, pGM297; organism, synthetic construct
 ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGG
 TTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTTGACGTCAATAATGACGT
 45 ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
 TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGC
 ATTATGCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
 TGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCATCTCCCCCCCCCTCCCCACCCCCAATTTTGTATTTAT
 TTATTTTTTAATTATTTTGTGCAGCATGGGGGCGGGGGGGGGGGGGGGGGGGCGCGCGCCAGCGGGGGGGGGGG
 50 GCGAGGGGCGGGGGCGGGGCGAGGCGGAGAGGTGCGGGCGGCAATCAGAGCGGCGCGCTCCGAAAGTTTTCCTT
 TTATGGCGAGGCGGGCGGGCGGGCGGCTATAAAAAAGCGAAGCGCGCGGGCGGGAGTTCGTGCGCGCTGCC
 TTCGCCCCGTGCCCGCTCCGCCGCCGCTCGCGCCCGCCCGCCCGGCTCTGACTGACCGCTTACTCCACAGG

TGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTCCTTTCTGT
GGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGAGCGGCTCGGGGGGTGCGTGCGTGT
GTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCCGCGGCTGTGAGCGCTGCGGGCGGGCGCGGGGC
TTTGTGCGCTCCGCGAGTGTGCGCGAGGGGAGCGCGGGCGGGGGCGGTGCCCCGCGGTGCGGGGGGGCTGCGAGG
5 GGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGTCGGTCGGGCTGCAAC
CCCCCTGCACCCCCCTCCCGAGTTGCTGAGCAGGCCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGCGG
CGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGTGGGGGTGCCGGGCGGGGCGGGGCCCTCGGGCCGGGG
AGGGCTCGGGGAGGGGCGCGGCGGCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGACGCCATTGCC
TTTTATGGTAAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGC
10 GCCGCCGCACCCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCC
TTCGTGCGTGCGCCGCGCCGCGCTCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCCTC
GGGGGGGACGGGGCAGGGCGGGGTTCCGGCTTCTGGCGTGTGACCGCGGGCTCTAGAGCCTCTGCTAACCATGTTC
ATGCCTTCTTCTTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAAAT
TGCTCGAGACTAGTGACTTGGTGAGTAGGCTTCGAGCCTAGTTAGAGGACTAGGAGAGGCCGTAGCCGTAACATAC
15 TCTGGGCAAGTAGGGCAGGCGGTGGGTACGCAATGGGGCGGGCTACCTCAGCACTAAATAGGAGACAATTAGACC
AATTTGAGAAAATACGACTTCGCCCCAACGGAAGAAAAAGTACCAAATTAACATTTAATATGGGCAGGCAAGG
AGATGGAGCGCTTCGGCCTCCATGAGAGGTTGTTGGAGACAGAGAGGGGTGTAAAAGAATCATAGAACTCCTCT
ACCCCCTAGAACCAACAGGATCGGAGGGCTTAAAAAGTCTGTTCAATCTTGTGTGCGTACTATATTGCTTGACACA
AGGAACAGAAAGTGAAGACACAGAGGAAGCAGTAGCAACAGTAAGACAACACTGCCATCTAGTGAAAAAGAAAA
20 AAAGTGCAACAGAGACATCTAGTGACAAAAAGAAAAATGACAAGGAATAGCAGCGCCACCTGGTGGCAGTCAGA
ATTTTCCAGCGCAACAACAAGGAAATGCCTGGGTACATGTACCCTTGTACCGCGCACCTTAAATGCGTGGGTAA
AAGCAGTAGAGGAGAAAAAATTTGGAGCAGAAATAGTACCCATGTTTCAAGCCCTATCAGAAGGCTGCACACCCCT
ATGACATTAATCAGATGCTTAATGTGCTAGGAGATCATCAAGGGGCATTACAAATAGTGAAAGAGATCATTAATG
AAGAAGCAGCCAGTGGGATGTAACACACCCACTACCCGCAGGACCCCTACCAGCAGGACAGCTCAGGGACCCCTC
25 GCGGCTCAGATATAGCAGGGACCACGCTCAGTACAAGAACAGTTAGAATGGATCTATACTGCTAACCCCCGGG
TAGATGTAGGTGCCATCTACCGGAGATGGATTATTTCTAGGACTTCAAAAGTGTGTCAAAATGTACAACCCAGTAT
CAGTCTTAGACATTAGGCAGGGACCTAAAGAGCCCTCAAGGATTATGTGGACAGATTTTACAAGGCAATTAGAG
CAGAACAAGCCTCAGGGGAAGTGAACAATGGATGACAGAATCATTACTCATTCAAAATGCTAATCCAGATTGTA
AGGTCATCCTGAAGGGCCTAGGAATGCACCCACCCCTTGAAGAAATGTTAACGGCTTGTACGGGGGTAGGAGGCC
30 CAAGCTACAAAGCAAAAGTAATGGCAGAAATGATGCAGACCATGCAAAATCAAAACATGGTGCAGCAGGGAGGTC
CAAAAAGACAAAGACCCCACTAAGATGTTATAATTGTGGAAAATTTGGCCATATGCAAAGACAATGTCCGGAAC
CAAGGAAAAACAAATGTCTAAAGTGTGGAAAATTTGGGACACCTAGCAAAAGACTGCAGGGGACAGGTGAATTTT
TAGGGTATGGACGGTGGATGGGGGCAAAACCGAGAAATTTCCCGCCGCTACTCTTGGAGCGGAACCGAGTGCGC
CTCCTCCACCGAGCGGCACCACCCCATACGACCCAGCAAGAAGCTCCTGCAGCAATATGCAGAGAAAGGGAAAC
35 AACTGAGGGAGCAAAAGAGGAATCCACCGCAATGAATCCGGATTGGACCGAGGGATATTCTTTGAACTCCCTCT
TTGGAGAAGACCAATAAAGACAGTGTATATAGAAGGGGTCCCCATTAAGGCACTGCTAGACACAGGGGCAGATGA
CACCATAATTAAGAAAAATGATTTACAATTATCAGGTCCATGGAGACCCAAAAATTATAGGGGCATAGGAGGAGG
CCTTAATGTAAAAGAAATATAACGACAGGGAAGTAAAAATAGAAGATAAAATTTTGAGAGGAACAATATTGTTAGG
AGCAACTCCCATTAATATAATAGGTAGAAATTTGTGCCCCGGCAGGTGCCCGGTTAGTAATGGGACAATTATC
40 AGAAAAAATTCCTGTCACACCTGTCAAATTGAAGGAAGGGGCTCGGGGACCCTGTGTAAGACAATGGCCTCTCTC
TAAAGAGAAGATTGAAGCTTTACAGGAAATATGTTCCCAATTAGAGCAGGAAGGAAAAATCAGTAGAGTAGGAGG
AGAAAAATGCATACAATACCCCAATATTTGCATAAAGAAGAAGGACAAATCCCAGTGGAGGATGCTAGTAGACTT
TAGAGAGTTAAATAAGGCAACCCCAAGATTTCTTTGAAGTGCAATTAGGGATACCCACCCAGCAGGATTAAGAAA
GATGAGACAGATAACAGTTTATAGATGTAGGAGACGCTATTATCCATACCATTGGATCCAAATTTTAGGAAATA
45 TACTGCTTTTACTATTCCACAGTGAATAATCAGGGACCCGGGATTAGGTATCAATTCAACTGTCTCCCGCAAGG
GTGGAAAGGATCTCCTACAATCTTCCAAAATACAGCAGCATCCATTTTGGAGGAGATAAAAAGAACTTGCCAGC
ACTAACCATTTGACAATACATGGATGATTTATGGGTAGGTTCTCAAGAAAATGAACACACCCATGACAAATTAGT
AGAACAGTTAAGAACAAAATTACAAGCCTGGGGCTTAGAAACCCAGAAAAGAAGGTGCAAAAAGAACCACCTTA
TGAGTGGATGGGATACAACTTTGGCCTCACAAATGGGAATAAGCAGAATACAACCTGGAGGAAAAAGATGAATG
50 GACTGTCAATTGACATCCAGAAGTTAGTTGGGAACTAAATTGGGCAGCACAATTGTATCCAGTCTTAGGACCAA
GAATATATGCAAGTTAATTAGAGGAAAGAAAAATCTGTTAGAGCTAGTGACTTGGACACCTGAGGCAGAAGCTGA
ATATGCAGAAAATGCAGAGATTCTTAAACAGAACGGAAGGAACCTATTACAAACCAGGAATACCTATTAGGGC
AGCAGTACAGAAATTTGAAGGAGGACAGTGGAGTTACCAATTCAAACAAGAAGGACAAGTCTTGAAAGTAGGAAA

ATACACCAAGCAAAAAGAACACCCATACAAATGAACTTCGCACATTAGCTGGTTTAGTGCAGAAGATTTGCAAAGA
AGCTCTAGTTATTTGGGGGATATTACCAGTTCTAGAACTCCCGATAGAAAGAGAGGTATGGGAACAATGGTGGGC
GGATTACTGGCAGGTAAGCTGGATTCCCGAATGGGATTTTGTTCAGCACCCACCTTTGCTCAAACATATGGTACAC
ATTAACAAAAGAACCATAACCAAGGAGGACGTTTACTATGTAGATGGAGCATGCAACAGAAATTCAAAAGAAGG
5 AAAAGCAGGATACATCTCACAATACGGAAAACAGAGAGTAGAAACATTAGAAAACACTACCAATCAGCAAGCAGA
ATTAACAGCTATAAAAAATGGCTTTGGAAGACAGTGGGCCTAATGTGAACATAGTAACAGACTCTCAATATGCAAT
GGGAATTTTGCAGCACAACCCACACAAAGTGATTACCATTTAGTAGAGCAAATATAGCCTTAATGATACAAAA
GCAACAAATATATTTGCAGTGGGTACCAGCACATAAAGGAATAGGAGGAAATGAGGAGATAGATAAAATTAGTGAG
TAAAGGCATTAGAAGAGTTTATTCTTAGAAAAATAGAGAAGCTCAAGAAGAGCATGAAAGATATCATATAA
10 TTGGAAAAACCTAGCAGATACATATGGGCTTCCACAAATAGTAGCAAAAGAGATAGTGGCCATGTGTCCAAAATG
TCAGATAAAGGGAGAACCAGTGCATGGACAAGTGGATGCCTCACCTGGAAACATGGCAGATGGATTGTACTCATCT
AGAAGGAAAAGTAGTCATAGTTGCGGTCCATGTAGCCAGTGGATTCTAGAGAAGCAGAGTCATACCTAGGGAAAC
AGGAAAAGAAACGGCAAAGTTTCTATTAATAAATACTGAGTAGATGGCCTATAACACAGTTACACACAGACAATGG
GCCTAACTTTACCTCCCAAGAAGTGGCAGCAATATGTTGGTGGGGAAAAATTGAACATACAACAGGTATACCATA
15 TAACCCCAATCTCAAGGATCAATAGAAAAGCATGAACAAACAATTAAGAGATAATTGGGAAAAAAGAGATGA
TTGCCAATATACAGAGACAGCAGTACTGATGGCTTGCCATATTACAATTTTAAAGAAAGGGAGGAATAGGGGG
ACAGACTTTCAGCAGAGAGACTAATTAATATAATAACAACACAATTAGAAATACAAACATTTACAAACCAAAATTC
AAAAATTTTAAATTTTAGAGTCTACTACAGAGAAGGAGAGACCCCTGTGTGGAAAGGACCAGCACAATTAATCTG
GAAAGGGGAAGGAGCAGTGGTCCTCAAGGACGGAAGTGACCTAAAGGTTGTACCAAGAAGGAAAGCTAAAAATTAT
20 TAAGGATTATGAACCCAAACAAAGAGTGGGTAATGAGGGTGACGTGGAAGGTACCAGGGGATCTGATAACTAAAT
GGCAGGGAATAGTCAGATATTGGATGAGACAAAGAAATTTGAAATGGAACATTTATATGCATCAGCTGGCGGCCG
CGAATTCAC TAGTGATTCCCGTTTGTGCTAGGGTTCTTAGGCTTCTTGGGGGCTGCTGGAACATGCAATGGGAGCA
GCGGCGACAGCCCTGACGGTCCAGTCTCAGCATTGCTTGTGGGATACTGCAGCAGCAGAAGAATCTGCTGGCG
GCTGTGGAGGCTCAACAGCAGATGTTGAAGCTGACCATTGCGGTGTTAAAAACCTCAATGCCCGCGTCACAGCC
25 CTGAGAAGTACCTAGAGGATCAGGCACGACTAACTCCTGGGGTGCGCATGGAACAAGTATGTCATACCACA
GTGGAGTGGCCCTGGACAAATCGGACTCCGATTGGCAAAATATGACTTGGTTGGAGTGGGAAAGACAAATAGCT
GATTTGGAAAGCAACATTACGAGACAATTAGTGAAGGCTAGAGAACAAAGAGGAAAGAATCTAGATGCCTATCAG
AAGTTAACTAGTTGGTCAGATTTCTGGTCTTGGTTCGATTTCTCAAATGGCTTAACATTTTAAATGGGATTT
TTAGTAATAGTAGGAATAATAGGGTTAAGATTACTTTACACAGTATATGGATGTATAGTGAGGGTTAGGCAGGGA
30 TATGTTCCCTCTATCTCCACAGATCCATATCCAATCGAATCCCGCGGCCGCAATTCACCTCCTCAGGTGCAGGCTG
CCTATCAGAAGGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCCCTCTGCCA
AAAAATTATGGGGACATCATGAAGCCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAA
TAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAAACATCAGAATG
AGTATTTGGTTTAGAGTTTGGCAACATATGCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTC
35 ATCAGTATATGAAACAGCCCCCTGCTGTCCATTCTTATTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTT
TTATATTTTGTGTTGTTATTTTTTCTTTAACATCCCTAAAATTTTCTTTACATGTTTTACTAGCCAGATTTT
TCCTCCTCTCCTGACTACTCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCAAGC
TTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAAATCCACACAACATACGAGCC
GGAAGCATAAAGTGTAAGCCTGGGGTGCTAATGATGAGCTAACTCACATTAATTGCGTTGGGCTCACTGCCC
40 GCTTTCCAGTCGGGAAACCTGTGCTGCCAGCGGATCCGCTATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAA
CTCCGCCCCATCCCGCCCCTAACTCCGCCCCAGTTCCGCCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTTATTT
ATGCAGAGGCCGAGGCCGCTCGGCCCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCCTAGGCT
TTTGCAAAAAGCTAACTTGTATTATGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAA
TAAAGCATTTTTTCTACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTCCGCT
45 TCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATA
CGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGT
AAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCT
GTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCA
50 CGCTGTAGGTATCTCAGTTCCGTTGATGTCGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC
GACCGCTGCGCCTTATCCGGTAACATTCGTCTTGTAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGTACAGAGTTCTTGAAGTGGTGGCCTAACTAC
GGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC

SEQ ID NO: 10 Exemplified hCEF promoter

	1	AGATCTGTTA	CATAACTTAT	GGTAAATGGC	CTGCCTGGCT	GACTGCCCAA	TGACCCCTGC
25	61	CCAATGATGT	CAATAATGAT	GTATGTTCCC	ATGTAATGCC	AATAGGGACT	TTCCATTGAT
	121	GTCAATGGGT	GGAGTATTTA	TGGTAACTGC	CCACTTGGCA	GTACATCAAG	TGTATCATAT
	181	GCCAAGTATG	CCCCCTATTG	ATGTCAATGA	TGGTAAATGG	CCTGCCTGGC	ATTATGCCCA
	241	GTACATGACC	TTATGGGACT	TTCCTACTTG	GCAGTACATC	TATGTATTAG	TCATTGCTAT
	301	TACCATGGGA	ATTCACTAGT	GGAGAAGAGC	ATGCTTGAGG	GCTGAGTGCC	CCTCAGTGGG
30	361	CAGAGAGCAC	ATGGCCCACA	GTCCCTGAGA	AGTTGGGGGG	AGGGGTGGGC	AATTGAACTG
	421	GTGCCTAGAG	AAGGTGGGGG	TTGGGTAAAC	TGGGAAAGTG	ATGTGGTGTA	CTGGCTCCAC
	481	CTTTTTCCCC	AGGGTGGGGG	AGAACCATAT	ATAAGTGAG	TAGTCTCTGT	GAACATTCAA
	541	GCTTCTGCCT	TCTCCCTCCT	GTGAGTTTGC	TAGC		

Length: 873; Molecule Type: DNA; Features Location/Qualifiers: source, 1..873; mol_type, unassigned DNA; organism, Human cytomegalovirus

CCGCGGAGATCTCAATATTGGCCATTAGCCATATATTTCATTGGTTATATAGCATAAATCAATATTGGCT
ATTGGCCATTGCATACGTTGTATCTATATCATAAATATGTACATTTATATTGGCTCATGTCCAATATGACC
40 GCCATTGTGGCATTGATTATTGACTAGTTATTAATAGTAACATTCACGGGTCATTAGTCTACATGCCCA
TATATGGAGTTCGCGGTTACATAAECTTACGGTAAATGGCCCGCTCGGCTGACCGCCCAACGACCCCCGCC
CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGT
GGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGATCATATGCCAAGTCCGCCCCCTATT
GACGTCATACGCGGTAATGGCCCCGCTGGCATTATGCCCGAGTACATGACCTTACGGGACTTTCCCTACTT
45 GGCAGTACATCTACGTATTAGTACTCGCTATTACCATGGTGAATCGGGTTTGGCAGTACACCAATGGGCGG
TGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTTTGG
CACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAATGGGCGGTAGGC
GTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCACTAGAAGCTTTATTGCG
GGTAGTTTATCACAGTTAAATTTGCTAACGCAGTCAGTGCTTCTGACACAACAGTCTCGAACTTAAGCTGC
50 AGAAGTTGGTCTGTGAGGCCTGGGCAGGCTAGC

SEQ ID NO: 12 Exemplified EF1a promoter

Length: 395; Molecule Type: DNA; Features Location/Qualifiers: source, 1..395; mol_type, unassigned DNA; organism, Homo sapiens

5 AGATCCATATCCGCGGCAATTTTAAAAGAAAGGGAGGAATAGGGGGACAGACTTCAGCAGAGAGACTAATTAATA
 TAATAACAACACAATTAGAAATACAACATTTACAAACCAAAATTCAAAAATTTTAAATTTTAGAGCCGCGGAGA
 TCCCGTGAGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCACAGTCCCCGAGAAGTTGGGGGGAGGGG
 TCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCC
 TTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGAACGGGTTTG
 CCGCCAGAACACAGGCTAGC

10

SEQ ID NO: 13 Exemplified CFTR transgene (soCFTR2)

Length: 4459; Molecule Type: DNA; Features Location/Qualifiers: source, 1..4459; mol_type, other DNA; note, soCFTR2; organism, synthetic construct

15 1 GCTAGCCACC ATGCAGAGAA GCCCTCTGGA GAAGGCCTCT GTGGTGAGCA AGCTGTTCTT
 61 CAGCTGGACC AGGCCCATCC TGAGGAAGGG CTACAGGCAG AGACTGGAGC TGTCTGACAT
 121 CTACCAGATC CCCTCTGTGG ACTCTGCTGA CAACCTGTCT GAGAAGCTGG AGAGGGAGTG
 181 GGATAGAGAG CTGGCCAGCA AGAAGAAGCC CAAGCTGATC AATGCCCTGA GGAGATGCTT
 241 CTTCTGGAGA TTCATGTTCT ATGGCATCTT CCTGTACCTG GGGGAAGTGA CCAAGGCTGT
 20 301 GCAGCCTCTG CTGCTGGGCA GAATCATTGC CAGCTATGAC CCTGACAACA AGGAGGAGAG
 361 GAGCATTGCC ATCTACCTGG GCATTGGCCT GTGCCTGCTG TTCATTGTGA GGACCTGTCT
 421 GCTGCACCCCT GCCATCTTGG GCCTGCACCA CATGGCATG CAGATGAGGA TTGCCATGTT
 481 CAGCCTGATC TACAAGAAAA CCCTGAAGCT GTCCAGCAGA GTGCTGGACA AGATCAGCAT
 541 TGGCCAGCTG GTGAGCCTGC TGAGCAACAA CCTGAACAAG TTTGATGAGG GCCTGGCCCT
 25 601 GGCCCACTTT GTGTGGATTG CCCCTCTGCA GGTGGCCCTG CTGATGGGCC TGATTTGGGA
 661 GCTGCTGCAG GCCTCTGCCT TTTGTGGCCT GGGCTTCCTG ATTGTGCTGG CCCTGTTTCA
 721 GGCTGGCCTG GGCAGGATGA TGATGAAGTA CAGGGACCAG AGGGCAGGCA AGATCAGTGA
 781 GAGGCTGGTG ATCACCTCTG AGATGATTGA GAACATCCAG TCTGTGAAGG CCTACTGTTG
 841 GGAGGAAGCT ATGGAGAAGA TGATTGAAAA CCTGAGGCAG ACAGAGCTGA AGCTGACCAG
 30 901 GAAGGCTGCC TATGTGAGAT ACTTCAACAG CTCTGCCTTC TTCTTCTCTG GCTTCTTTGT
 961 GGTGTTCCCTG TCTGTGCTGC CCTATGCCCT GATCAAGGGG ATCATCCTGA GAAAGATTTT
 1021 CACCACCATC AGCTTCTGCA TTGTGCTGAG GATGGCTGTG ACCAGACAGT TCCCCTGGGC
 1081 TGTGCAGACC TGGTATGACA GCCTGGGGGC CATCAACAAG ATCCAGGACT TCCTGCAGAA
 1141 GCAGGAGTAC AAGACCTGG AGTACAACCT GACCACCACA GAAGTGGTGA TGGAGAATGT
 35 1201 GACAGCCTTC TGGGAGGAGG GCTTTGGGGA GCTGTTTGAG AAGGCCAAGC AGAACAACAA
 1261 CAACAGAAAG ACCAGCAATG GGGATGACTC CCTGTTCTTC TCCAACCTCT CCCTGCTGGG
 1321 CACACCTGTG CTGAAGGACA TCAACTTCAA GATTGAGAGG GGGCAGCTGC TGGCTGTGGC
 1381 TGGATCTACA GGGGCTGGCA AGACCAGCCT GCTGATGATG ATCATGGGGG AGCTGGAGCC
 1441 TTCTGAGGGC AAGATCAAGC ACTCTGGCAG GATCAGCTTT TGCAGCCAGT TCAGCTGGAT
 40 1501 CATGCCTGGC ACCATCAAGG AGAACATCAT CTTTGGAGTG AGCTATGATG AGTACAGATA
 1561 CAGGAGTGTG ATCAAGGCCT GCCAGCTGGA GGAGGACATC AGCAAGTTTG CTGAGAAGGA
 1621 CAACATTGTG CTGGGGGAGG GAGGCATTAC ACTGTCTGGG GGCCAGAGAG CCAGAATCAG
 1681 CCTGGCCAGG GCTGTGTACA AGGATGCTGA CCTGTACCTG CTGGACTCCC CCTTTGGCTA
 1741 CCTGGATGTG CTGACAGAGA AGGAGATTTT TGAGAGCTGT GTGTGCAAGC TGATGGCCAA
 45 1801 CAAGACCAGA ATCCTGTGTA CCAGCAAGT GGAGCACCTG AAGAAGGCTG ACAAGATCCT
 1861 GATCCTGCAT GAGGGCAGCA GCTACTTCTA TGGGACCTTC TCTGAGCTGC AGAACCTGCA
 1921 GCCTGACTTC AGCTCTAAGC TGATGGGCTG TGACAGCTTT GACCAGTTCT CTGCTGAGAG
 1981 GAGGAACAGC ATCCTGACAG AGACCCTGCA CAGATTCAGC CTGGAGGGAG ATGCCCTGT
 2041 GAGCTGGACA GAGACCAAGA AGCAGAGCTT CAAGCAGACA GGGGAGTTTG GGGAGAAGAG
 50 2101 GAAGAACTCC ATCCTGAACC CCATCAACAG CATCAGGAAG TTCAGCATTG TGCAGAAAAA
 2161 CCCCCTGCAG ATGAATGGCA TTGAGGAAGA TTCTGATGAG CCCCTGGAGA GGAGACTGAG
 2221 CCTGGTGCCT GATTCTGAGC AGGGAGAGGC CATCCTGCCT AGGATCTCTG TGATCAGCAC
 2281 AGGCCCTACA CTGCAGGCCA GAAGGAGGCA GTCTGTGCTG AACCTGATGA CCCACTCTGT
 2341 GAACCAGGGC CAGAACATCC ACAGGAAAAC CACAGCCTCC ACCAGGAAAG TGAGCCTGGC
 55 2401 CCCTCAGGCC AATCTGACAG AGCTGGACAT CTACAGCAGG AGGCTGTCTC AGGAGACAGG
 2461 CCTGGAGATT TCTGAGGAGA TCAATGAGGA GGACCTGAAA GAGTGCTTCT TTGATGACAT
 2521 GGAGAGCATC CCTGCTGTGA CCACCTGGAA CACCTACCTG AGATACATCA CAGTGCACAA

2581 GAGCCTGATC TTTGTGCTGA TCTGGTGCCT GGTGATCTTC CTGGCTGAAG TGGCTGCCTC
2641 TCTGGTGGTG CTGTGGCTGC TGGGAAACAC CCCACTGCAG GACAAGGGCA ACAGCACCCA
2701 CAGCAGGAAC AACAGCTATG CTGTGATCAT CACCTCCACC TCCAGCTACT ATGTGTTCTA
2761 CATCTATGTG GGAGTGGCTG ATACCCTGCT GGCTATGGGC TTCTTTAGAG GCCTGCCCTC
5 2821 GGTGCACACA CTGATCACAG TGAGCAAGAT CCTCCACCAC AAGATGCTGC ACTCTGTGCT
2881 GCAGGCTCCT ATGAGCACCC TGAATACCCT GAAGGCTGGG GGCATCCTGA ACAGATTCTC
2941 CAAGGATATT GCCATCCTGG ATGACCTGCT GCCTCTCACC ATCTTTGACT TCATCCAGCT
3001 GCTGCTGATT GTGATTGGGG CCATTGCTGT GGTGGCAGTG CTGCAGCCCT ACATCTTTGT
3061 GGCCACAGTG CCTGTGATTG TGGCCTTCAT CATGCTGAGG GCCTACTTTC TGCAGACCTC
10 3121 CCAGCAGCTG AAGCAGCTGG AGTCTGAGGG CAGAAGCCCC ATCTTCACCC ACCTGGTGAC
3181 AAGCCTGAAG GGCCTGTGGA CCCTGAGAGC CTTTGGCAGG CAGCCCTACT TTGAGACCCT
3241 GTTCCACAAG GCCCTGAACC TGCACACAGC CAACTGGTTC CTCTACCTGT CCACCCTGAG
3301 ATGGTTCCAG ATGAGAATTG AGATGATCTT TGTGATCTTC TTCATTGCTG TGACCTTCAT
3361 CAGCATTCTG ACCACAGGAG AGGGAGAGGG CAGAGTGGGC ATTATCCTGA CCCTGGCCAT
15 3421 GAACATCATG AGCACACTGC AGTGGGCAGT GAACAGCAGC ATTGATGTGG ACAGCCTGAT
3481 GAGGAGTGTG AGCAGAGTGT TCAAGTTCAT TGATATGCCC ACAGAGGGCA AGCCTACCAA
3541 GAGCACCAG CCCTACAAGA ATGGCCAGCT GAGCAAAGTG ATGATCATTG AGAACAGCCA
3601 TGTGAAGAA GATGATATCT GGCCAGTGG AGGCCAGATG ACAGTGAAGG ACCTGCACAGC
3661 CAAGTACACA GAGGGGGGCA ATGCTATCCT GGAGAACATC TCCTTCAGCA TCTCCCTGG
20 3721 CCAGAGAGTG GGAAGTGTGG GAAGAACAGG CTCTGGCAAG TCTACCCTGC TGTCTGCCTT
3781 CCTGAGGCTG CTGAACACAG AGGGAGAGAT CCAGATTGAT GGAGTGTCTT GGGACAGCAT
3841 CACACTGCAG CAGTGGAGGA AGGCCTTTGG TGTGATCCCC CAGAAAGTGT TCATCTTCAG
3901 TGGCACCTTC AGGAAGAACC TGGACCCCTA TGAGCAGTGG TCTGACCAGG AGATTTGGAA
3961 AGTGGCTGAT GAAGTGGGCC TGAGAAGTGT GATTGAGCAG TTCCCTGGCA AGCTGGACTT
25 4021 TGTCTGTGGT GATGGGGGCT GTGTGCTGAG CCATGGCCAC AAGCAGCTGA TGTGCTGGC
4081 CAGATCAGTG CTGAGCAAGG CCAAGATCCT GCTGCTGGAT GAGCCTTCTG CCCACCTGGA
4141 TCCTGTGACC TACCAGATCA TCAGGAGGAC CCTCAAGCAG GCCTTTGCTG ACTGCACAGT
4201 CATCCTGTGT GAGCACAGGA TTGAGGCCAT GCTGGAGTGC CAGCAGTTCC TGGTGAATGA
4261 GGAGAACAAA GTGAGGCAGT ATGACAGCAT CCAGAAGCTG CTGAATGAGA GGAGCCTGTT
30 4321 CAGGCAGGCC ATCAGCCCTT CTGATAGAGT GAAGCTGTTC CCCCACAGGA ACAGCTCCAA
4381 GTGCAAGAGC AAGCCCCAGA TTGCTGCCCT GAAGGAGGAG ACAGAGGAGG AAGTGCAGGA
4441 CACCAGGCTG TGAGGGCCC

35 SEQ ID NO: 14 Exemplified A1AT transgene

Length: 1257; Molecule Type: DNA; Features Location/Qualifiers: source, 1..1257; mol_type, other DNA; note, sohAAT organism, synthetic construct

ATGCCAGCTCTGTGTCCTGGGGCATTCTGCTGCTGGCTGGCCTGTGCTGTCTGGTGCCTGTGTCCCTGG
40 CTGAGGACCCTCAGGGGATGCTGCCCAGAAAACAGACACCTCCACCATGACCAGGACCACCCACCTT
CAACAAGATCACCCCAACCTGGCAGAGTTTGCCCTTCAGCCTGTACAGACAGCTGGCCACACAGAGCAAC
AGCACCAACATCTTTTTACGCCCTGTGTCCATTGCCACAGCCTTTGCCATGCTGAGCCTGGGCACCAAGG
CTGACACCCATGATGAGATCCTGGAAGGCCTGAACCTTCAACCTGACAGAGATCCCTGAGGCCAGATCCA
TGAGGGCTTCCAGGAACCTGCTGAGAACCCTGAACCAGCCAGACAGCCAGCTGCAGCTGACAACAGGCAAT
45 GGGCTGTTCTGTCTGAGGGCCTGAAGCTGGTGGACAAGTTTCTGGAAGATGTGAAGAAGCTGTACCACT
CTGAGGCCTTCACAGTGAACCTTTGGGGACACAGAAGAGGCCAAGAAACAGATCAATGACTATGTGAAAAA
GGGCACCCAGGGCAAGTGTGGACCTTGTGAAAGCTGGACAGGGACACTGTGTTTGCCTTGTGAAC
TACATCTTCTTCAAGGCAAGTGGGAGAGGCCCTTTGAAGTGAAGGACACTGAGGAAGAGGACTTCCATG
TGGACCAAGTGACCACAGTGAAGGTGCCAATGATGAAGAGACTGGGGATGTTCAATATCCAGCACTGCAA
50 GAAACTGAGCAGCTGGGTGCTGCTGATGAAGTACCTGGGCAATGCTACAGCCATATTCTTTCTGCCTGAT
GAGGGCAAGCTGCAGCACCTGGAAAATGAGCTGACCCATGACATCATACCAAATTTCTGGAAAATGAGG
ACAGAAGATCTGCCAGCCTGCATCTGCCCAGCTGAGCATCACAGGCACATATGACCTGAAGTCTGTGCT
GGGACAGCTGGGAATCACCAAGGTGTTAGCAATGGGGCAGACCTGAGTGGAGTGACAGAGGAAGCCCCT
CTGAAGCTGTCCAAGCTGTGCACAAGGCAGTGTGACCATGATGAGAAGGGCACAGAGGCTGTGGGG
55 CCATGTTTCTGGAAGCCATCCCCATGTCCATCCCCCAGAAGTGAAGTTCAACAAGCCCTTGTGTTCCT
GATGATTGAGCAGAACCAAGAGCCCCCTGTTATGGGCAAGGTTGTGAACCCACCCAGAAATGA

SEQ ID NO: 15 Complementary strand to the exemplified A1AT transgene

Length: 1257; Molecule Type: DNA; Features Location/Qualifiers: source, 1..1257; mol_type, other DNA; note, sohAAT complementary strand;

5 organism, synthetic construct

TACGGGTCGAGACACAGGACCCCGTAAGACGACGACCGACCGGACACGACAGACCACGGACACAGGGACC
 GACTCCTGGGAGTCCCCCTACGACGGGTCTTTTGTCTGTGGAGGGTGGTACTGGTCCTGGTGGGGTGGAA
 GTTGTCTAGTGGGGGTGGACCGTCTCAAACGGAAGTCGGACATGTCTGTCGACCGGGTGGTCTCGTTG
 TCGTGGTTGTAGAAAAAGTCGGGACACAGGTAACGGTGTGCGAAACGGTACGACTCGGACCCGTGGTTCC
 10 GACTGTGGGTACTACTCTAGGACCTTCCGGACTTGAAGTTGGACTGTCTCTAGGGACTCCGGGTCTAGGT
 ACTCCCGAAGGTCCTTGACGACTCTTGGGACTTGGTCGGTCTGTGCGTCGACGTCGACTGTGTCCGTTA
 CCCGACAAGGACAGACTCCCGGACTTCGACCACCTGTTCAAAGACCTTCTACACTTCTTCGACATGGTGA
 GACTCCGGAAGTGTCACTTGAAACCCCTGTGTCTTCTCCGGTCTTTGTCTAGTTACTGATACACCTTTT
 CCCGTGGGTCCCGTTCTAACACCTGGAACACTTCTCGACCTGTCCCTGTGACACAAACGGGAACACTTG
 15 ATGTAGAAGAAGTTCCTCCGTTACCCCTCTCCGGGAACTTCACTTCCTGTGACTCCTTCTCCTGAAGGTAC
 ACCTGGTTCACTGGTGTCACTTCCACGGTTACTTCTCTGACCCCTACAAGTTATAGGTCTGTGACGTT
 CTTTGACTCGTCGACCCACGACGACTACTTCACTGGACCCGTTACGATGTGCGGTATAAGAAAGACGGACTA
 CTCCCGTTCGACGTCGTGGACCTTTTACTCGACTGGGTACTGTAGTAGTGGTTTAAAGACCTTTTACTCC
 TGTCTTCTAGACGGTCGGACGTAGACGGGTTGACTCGTAGTGTCCGTGTATACTGGACTTCAGACACGA
 20 CCTGTGCGACCTTAGTGGTTCACAAAGTCGTACCCCGTCTGGACTCACCTCACTGTCTCCTTCGGGGA
 GACTTCGACAGGTTCCGACACGTGTTCCGTACGACTGGTAACTACTCTTCCCGTGTCTCCGACGACCCC
 GGTACAAAGACCTTCGGTAGGGGTACAGGTAGGGGGTCTTCACTTCAAGTTGTTTCGGGAAACACAAGGA
 CTACTAACTCGTCTTGTGGTCTCTCGGGGACAAAGTACCCGTTCCAACACTTGGGGTGGGTCTTTACT

25

SEQ ID NO: 16 Exemplified A1AT polypeptide

Length: 419; Molecule Type: AA; Features Location/Qualifiers: SOURCE, 1..419; MOL_TYPE, protein; ORGANISM, Homo sapiens

AEDPQGDAQKTDTSHHQDHPFAEDPQGDAAQKTDTSHHQDHPFNKITPNLAEFAFSLYRQLAHQSN
 30 STNIFFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTETPEAQIHEGFQELLRLTNQPDSQLQLTTGNG
 LFLSEGLKLVDFKLEVDVKLYHSEAFVNFVGDTEEAQKQINDYVEKGTQGKIVDLVKELDRDRTVFALVNYI
 FFKGKWERPFVVKDTEEDFHVQVTTVKVPMKRLGMFNIHQCKKLSSWVLLMKYLGNAIAIFFLPDEGK
 LQHLENELTHDIIITKFLNEDRRSASLHLPKLSITGTYDLKSVLGLGITKVFNSGADLSGVTEEAPLKLS
 KAVHKAVLTIIDEKGTAAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK

35

SEQ ID NO: 17 Exemplified FVIII transgene (N6)

Length: 5013; Molecule Type: DNA; Features Location/Qualifiers: source, 1..5013; mol_type, other DNA; note, codon-optimised FVIII transgene

40 (N6); organism, synthetic construct

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGTGCCTGTGAGGTTCTGCTTCTCTGCCACCAGGAGAT
 ACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCTGTGGATGC
 CAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCTTCAACACCTCTGTGGTGTACAAGAAGACCTGTTT
 GTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCCAGGCCCCCTGGATGGGCTGTGGGCCCCA
 45 CCATCCAGGCTGAGGTGTATGACACTGTGGTGATCACCTGAAGAACATGGCCAGCCACCTGTGAGCCT
 GCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGG
 GAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCGACACACCTATGTGTGGCAGGTGCTGAAGGAGAATG
 GCCCATGGCCTCTGACCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCT
 GAACCTGTGGCTGATTGGGGCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACC
 50 CTGCACAAGTTATCCTGCTGTTTGTGTGTTTGTGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACA
 GCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGT
 GAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTGATTGGCATGGGC
 ACCACCCCTGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTGGTCAGGAACCACAGGCAGGCCA

70

GCCTGGACCCCCCCTGCTGACCAGATACCTGAGGATTACCCCCAGAGCTGGGTGCACCAGATTGCCCT
GAGGATGGAGGTGCTGGGCTGTGAGGCCAGGACCTGTACTGA

5 **SEQ ID NO: 18 Exemplified FVIII transgene (V3)**

Length: 4425; Molecule Type: DNA; Features Location/Qualifiers: source,
1..4425; mol_type, other DNA; note, codon-optimised FVIII transgene
(V3); organism, synthetic construct

10 ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGAT
ACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCGTGTGGATGC
CAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCTTCAACACCTCTGTGGTGTACAAGAAGACCCTGTTT
GTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCCAGGCCCCCTGGATGGGCCTGCTGGGCCCCA
CCATCCAGGCTGAGGTGTATGACACTGTGGTGATCACCTGAAGAACATGGCCAGCCACCCTGTGAGCCT
GCATGCTGTGGGGGTGAGCTACTGGAAGGCCCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGG
15 GAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCGACCCACACCTATGTGTGGCAGGTGCTGAAGGAGAATG
GCCCCATGGCCTCTGACCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCT
GAACTCTGGCCTGATTGGGGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACC
CTGCACAAGTTTCATCCTGCTGTTTGCTGTGTTGATGAGGGCAAGAGCTGGCACTCTGAACCAAGAACA
GCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCCAAGATGCACACTGTGAATGGCTATGT
20 GAACAGGAGCCTGCCITGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTGATTGGCATGGGC
ACCACCCCTGAGGTGCACAGCATCTTCTGGAGGGCCACACCTTCTGGTCAGGAACACAGGCAGGCCA
GCCTGGAGATCAGCCCCATCACCTTCTGACTGCCAGACCCTGCTGATGGACCTGGGCCAGTTCTCTGCT
GTTCTGCCACATCAGCAGCCACCAGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAG
GAGCCCCAGCTGAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGA
25 TGGATGTGGTGAGGTTTGTATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCA
CCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCTGGTGCTGGCC
CCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAAGA
AGGTCAGGTTTCATGGCCTACACTGATGAACTTCAAGACCAGGGAGGCCATCCAGCATGAGTCTGGCAT
CCTGGGCCCCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATCTTCAAGAACCAGGCCAGCAGG
30 CCCTACAACATCTACCCCCATGGCATCACTGATGTGAGGCCCTGTACAGCAGGAGGCTGCCAAGGGGG
TGAAGCACCTGAAGGACTTCCCCATCTGCCTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGA
GGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGG
GACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGGCAACC
AGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGTGATGAGAACAGGAGCTGGTACCTGAC
35 TGAGAACATCCAGAGGTTTCTGCCAACCCCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGC
AACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAGCTGTCTGTGTGCTGCATGAGGTGG
CCTACTGGTACATCCTGAGCATTGGGGCCGAGCTGACTTCTGCTGTGTTCTTCTCTGCTGATCACCTT
CAAGCACAAGATGGTGTATGAGGACACCCTGACCCTGTTCCCCTTCTCTGGGGAGACTGTGTTTCATGAGC
ATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAACCTGACTTCAGGAACAGGGGCATGACTGCC
40 TGCTGAAAGTCTCCAGCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGC
CTACCTGCTGAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATGCCACTAATGTGTCTAAC
AACAGCAACACCAGCAATGACAGCAATGTGTCTCCCCAGTGCTGAAGAGGCACCAGAGGGAGATCACCA
GGACCACCCTGCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAAGAAGGA
GGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACAGGCACCTAC
45 TTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCCCCATGTGCTGAGGAACAGGG
CCCAGTCTGGCTCTGTGCCCCAGTTCAAGAAGGTGGTGTTCAGGAGTTCACTGATGGCAGCTTCACCCA
GCCCCGTGTACAGAGGGGAGCTGAATGAGCACCTGGGCCTGCTGGGCCCCCTACATCAGGGCTGAGGTGGAG
GACAACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCT
ATGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCCCAATGAAACCAAGACCTA
50 CTTCTGGAAGGTGCAGCACCATGGCCCCCACAAGGATGAGTTTACTGCAAGGCCTGGGCTACTTCT
TCTGATGTGGACCTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCCCTGCTGGTGTGCCACACCAACA
CCCTGAACCCTGCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGCCTGTTCTTACCAATCTTTGATGA
AACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAAGTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGCC
55 TGGTGTGAGGCCAGGACCAGAGGATCAGGTGGTACCTGCTGAGCATGGGCAGCAATGAGAACATCCACAG
CATCCACTTCTCTGGCCATGTGTTCACTGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTG

10

20 Length: 5013; Molecule Type: DNA; Features Location/Qualifiers: source,
1..5013; mol_type, other DNA; note, codon-optimised FVIII transgene
(N6) complementary strand; organism, synthetic construct

50

TGGCCTGGGGGTACGGGTTCTAGGTCTTACACTCGTCGAGACTGGACGACTACGACGACTCCGTCTCGGG
 GTGGGGGTACCGGACTCGGACAGACTGGACGTCCTCCGGTTCATACTTTGGAAGAGACTACTGGGGTCG
 GGACCCCGGTAAGTGTCTGTTGTTCGGACAGACTCTACTGGGTGAAGTCCGGGGTTCGACGTGGTGAGAC
 CCTGTACACAAAGTGGGACTCAGACCGGACGTCGACTCCGACTTACTCTTCGACCCGTGGTGACGACG
 5 GTGACTCGACTTCTTCGACCTGAAGTTTCAGAGGTCTGTTGGTCTGTTGGTGGTACTAGTCTGGTAGGGGAGA
 CTGTTGGACCGACGACCGTGACTGTTGTGGTCTGTCGGACCCGGGGGGTCTGACGGACACGTGATACTGT
 CGGTTCGACCTGTGGTGGGACAAACCGTTCTTCTCGTCGGGGGACTGACTCAGACCCCGGGGGACTCGGA
 CAGACTCCTCTTGTACTGTCTGCTTCGACGACCTCAGACCGGACTACTTGTCTGGTCTCTCGTCGACCCCG
 TTCTTACACTCGTCGTCCTCTAGTGGTCTGGTGGGACGTCAGACTGGTCTCTCTAACTGATACTAC
 10 TGTGGTAGAGACACCTCTACTTCTTCTCTGAAACTGTAGATGCTGCTCTGCTCTTGGTCTCGGGGTC
 CTCGAAGGTCTTCTTCTGGTCCGTGATGAAGTAACGACGACACCTCTCCGACACCCTGATACCGTACTCG
 TCGTCGGGGGTACACGACTCCTTGTCCCGGTTCAGACCGAGACACGGGGTCAAGTTCTTCCACCACAAGG
 TCCTCAAGTGACTACCGTCGAAGTGGGTCTGGGACATGTCTCCCTCGACTTACTCGTGGACCCGGACGA
 CCCGGGGATGTAGTCCCGACTCCACCTCTGTTGTAGTACCCTGGAAGTCTTGGTCCGGTCTGTCGGG
 15 ATGTCGAAGATGTCGTCGACTAGTCGATACCTCCTCGTCTGCTCCGTCCCGGACTCGGGTCTTCTTGA
 AACACTTCGGGTTACTTTGGTCTGATGAAGACCTTCCACGTCGTGGTGTACCGGGGTGGTCTCTACT
 CAAACTGACGTTCCGGACCCGGATGAAGAGACTACACCTGGACCTCTTCTACACGTGAGACCGGACTAA
 CCGGGGGACGACACACGTTGTGGTGTGGGACTTGGGACGGGTACCGTCCGTCCACTGACACGTCCTCA
 AACGGGACAAGAAGTGGTAGAACTACTTTGGTCTCGACCATGAAGTGACTCTTGTACCTCTCCTTGAC
 20 GTCCCGGGGACGTTGTAGGTCTACCTCCTGGGGTGAAGTTCCTCTTGATGTCCAAGGTACGGTAGTTA
 CCGATGTAGTACCTGTGGGACGGACCGGACCACTACCGGGTCTGGTCTCTAGTCCACCATGGACGACT
 CGTACCCGTCGTTACTCTTGTAGGTGTCTAGGTGAAGAGACCGGTACACAAGTGACACTCCTTCTTCTCT
 CCTCATGTTCTACCGGGACATGTTGGACATGGGACCCACAAACTCTGACACCTCTACGACGGGTCTGTT
 CGACCGTAGACCTCCACCTCACGACTAACCCCTCGTGGACGTACGACCGTACTCGTGGGACAAGGACC
 25 ACATGTCGTTGTTACGGTCTGGGGGACCCGTACCGGAGACCGGTGTAGTCCCTGAAGGTCTAGTGACG
 GAGACCGGTTCATACCGGTACCCCGGGGTTTCGACCGGTCCGACGTGATGAGACCGTCTGAGTTACGGACC
 TCGTGGTTCCTCGGGAAGTCGACCTAGTTCACCTGGACGACCGGGGTACTAGTAGGTACCGTAGTTCT
 GGGTCCCCCGGTCCGTCTTCAAGTCGTCGGACATGTAGTCCGTCAAGTAGTAGTACATGTCGGACCTACC
 GTTCTTACCGTCTGGATGTCCCCGTTGTCTGACCGTGGGACTACCACAAGAAACCGTTACACCTGTCTG
 30 AGACCGTAGTTCGTGTTGTAGAAGTTGGGGGGGTAGTAACGGTCTATGTAGTCCGACGTGGGGTGGGTGA
 TGTCCGTAGTCTCGTGGGACTCCTACCTCGACTACCCGACACTGGACTTGTCTGACGTCTGACGGGACCC
 GTACCTCTCGTTCGGTAGAGACTACGGGTCTAGTGACGGTCTGATGAAGTGGTGTACAAACGGTGG
 ACCTCGGGGTCTGTTCCGGTCCGACGTGGACGTCCCGTCTCTGTTACGGACCTCCGGGGTCCAGTTGTTGG
 GGTCTCTACCGACGTCCACCTGAAGGTCTTCTGGTACTTCCACTGACCCCACTGGTGGGTCCCCCACTT
 35 CTCGGACGACTGGTCTGATACATACTTCTCAAGGACTAGTCGTCTGTCGGTCTACCGGTGGTCACTGG
 GACAAGAAGGTCTTACCGTTCCACTTCCACAAGGTCCCCTTGGTCTGTGCAAGTGGGGACACCACTTGT
 CGGACCTGGGGGGGACGACTGGTCTATGGACTCCTAAGTGGGGGTCTCGACCCACGTGGTCTAACGGGA
 CTCTACCTCCACGACCCGACACTCCGGGTCTGGACATGACT

40

SEQ ID NO: 20 Complementary strand to the exemplified FVIII transgene (V3)

Length: 4425; Molecule Type: DNA; Features Location/Qualifiers: source, 1..4425; mol_type, other DNA; note, codon-optimised FVIII transgene (V3) complementary strand; organism, synthetic construct

45 TACGTCTAACTCGACTCGTGGACGAAGAAGGACACGGACGACTCCAAGACGAAGAGACGGTGGTCTCTA
 TGATGGACCCCCGACACCTCGACTCGACCTGATGTACGTCAGACTGGACCCCTCGACGGACACCTACG
 GTCCAAGGGGGGGTCTACGGGTTCTCGAAGGGGAAGTTGTGGAGACACCACATGTTCTTCTGGGACAAA
 CACCTCAAGTGACTGGTGGACAAGTTGTAACGGTTCGGGTCCGGGGGACCTACCCGGACGACCCGGGGT
 50 GGTAGGTCCGACTCCACATACTGTGACACCACTAGTGGGACTTCTTGTACCGGTCCGTGGGACACTCGGA
 CGTACGACACCCCACTCGATGACCTTCCGGAGACTCCCCGACTCATACTACTGGTCTGGTCTGCTCTCC
 CTCTTCTCTACTGTTCCACAAGGGACCCCGTCCGTGTGGATACACACCGTCCACGACTTCTCTTAC
 CGGGGTACCGGAGACTGGGGGACACGGACTGGATGTCGATGGACTCGGTACACCTGGACCACTTCTGGA
 CTTGAGACCGGACTAACCCCGGGACGACACACGTCCCTCCCGTCCGACCGGTTCTCTTCTGGGTCTGG
 GACGTGTTCAAGTAGGACGACAAACGACACAACTACTCCCGTTCTCGACCGTGAGACTTTGGTCTTGT
 55 CGGACTACGTCCTGTCCCTACGACGGAGACGGTCCCGGACCGGGTCTACGTGTGACACTTACCGATACA
 CTTGTCTCTCGGACGGACCGGACTAACCGACGGTGTCTTCTCAGACACATGACCGTACACTAACCGTACCCG

TGGTGGGGACTCCACGTGTCGTAGAAGGACCTCCCGGTGTGGAAGGACCAGTCCTTGGTGTCCGTCCGGT
CGGACCTCTAGTCGGGGTAGTGGAAGGACTGACGGGTCTGGGACGACTACCTGGACCCGGTCAAGGACGA
CAAGACGGTGTAGTCGTGGTGGTCTGACTACCGTACCTCCGGATACACTTCCACCTGTCGACGGGACTC
CTCGGGGTCGACTCCTACTTCTTGTACTCCTCCGACTCCTGATACTACTACTGGACTGACTGAGACTCT
5 ACCTACACCACTCCAACTACTACTGTTGTGCGGGTCTGAAGTAGGTCTAGTCCAGACACCGGTTCTTCGT
GGGGTCTTGGACCCACGTGATGTAACGACGACTCCTCCTCCTGACCCTGATACGGGGGACACGACCGG
GGACTACTGTCTCGATGTTCTCGGTCTAGGACTTGTACCGGGGGTCTCCTAACCGTCTTCATGTTCT
TCCAGTCCAAGTACCGGATGTGACTACTTTGGAAGTTCTGGTCCCTCCGGTAGGTCTGACTCAGACCGTA
10 GGACCCGGGGACGACATACCCCTCCACCCCTGTGGGACGACTAGTAGAAGTTCTTGGTCCGGTCTGTCC
GGGATGTTGTAGATGGGGTACCGTAGTGACTACACTCCGGGGACATGTGCTCCTCCGACGGGTTCCCCC
ACTTCGTGGACTTCCCTGAAGGGGTAGGACGGACCCCTCTAGAAGTTCATGTTACCTGACACTGACACCT
CCTACCGGGGTGGTTTACACTGAGTGGGGTCCACGACTGGTCTATGATGTCGTCGAAACACTTGTACCTCTCC
CTGGACCGGAGACCGGACTAACCGGGGGACGACTAGACGATGTTCTCAGACACCTGGTCTCCCCGTTGG
TCTAGTACAGACTGTTCTCCTTACACTAGGACAAGAGACACAACTACTCTTGTCTCGACCATGGACTG
15 ACTCTTGTAGGTCTCCAAGGACGGGTTGGGACGACCCACGTCGACCTCCTGGGACTCAAGTCCGGTCTG
TTGTAGTACGTGTCGTAGTTACCGATACACAACTGTGCGACGTGACAGACACACGGACGTACTCCACC
GGATGACCATGTAGGACTCGTAACCCCGGGTCTGACTGAAGGACAGACACAAGAAGAGACCGATGTGGAA
GTTCTGTGTTCTACACATACTCCTGTGGGACTGGGACAAGGGGAAGAGACCCCTCTGACACAAGTACTCG
TACCTCTTGGGACCGGACACCTAAGACCCGACGGTGTGAGACTGAAGTCTTGTCCCGTACTGACGGG
20 ACGACTTTCAGAGGTCGACACTGTTCTTGTGACCCCTGATGATACTCCTGTGCGATACTCCTGTAGAGACG
GATGGACGACTCGTCTTGTACGGTAACTCGGGTCTCGAAGTCGGTCTTACGGTGATTACACAGATTG
TTGTCGTTGTGGTCTTACTGTGCTTACACAGAGGGGTCACGACTTCTCCGTGGTCTCCCTCTAGTGGT
CCTGGTGGGACGTACAGACTGGTCTCCTCTAACTGATACTACTGTGGTAGAGACACCTTACTTCTTCTCCT
CCTGAACTGTAGATGCTGCTCCTGCTCTTGGTCTCGGGGTCCTCGAAGGTCTTCTTGGTCCGTGATG
25 AAGTAACGACGACACCTCTCCGACACCCCTGATACCGTACTCGTCGTCGGGGGTACACGACTCCTTGTCCC
GGGTGACACCGAGACACGGGTCAAGTTCTTCCACCACAAGGTCCTCAAGTACTACCGTCAAGTGGGT
CGGGGACATGTCTCCCTCGACTTACTCGTGACCCCGGACGACCCGGGGATGTAGTCCCGACTCCACCTC
CTGTTGTAGTACCACTGGAAGTCCTTGGTCCGGTCTGTCGGAAGATGTGTCGCGACTAGTCTGA
TACTCCTCCTGGTCTCCGTCCCCGACTCGGGTCTTCTTGAACACTTCGGGTACTTTGGTCTGGAT
30 GAAGACCTTCCACGTGCTGGTGTACCGGGGGTGGTTCCTACTCAAAGTACGTTCCGGACCCGGATGAAG
AGACTACACCTGGACCTCTTCTTACACGTGAGACCGGACTAACCGGGGGACGACACACCGTGTGGTTGT
GGGACTTGGGACGGGTACCGTCCGTCCACTGACACGTCTCAAACGGGACAAGAAGTGGTAGAACTACT
TTGGTCTCGACCATGAAGTACTCTTGTACCTCTCCTTGACGTCCCGGGGGACGTTGTAGGTCTACCTC
CTGGGGTGAAGTCTCTTGTGATGTCCAAGTACGGTAGTTACCGATGTAGTACCTGTGGGACGGACCGG
35 ACCACTACCGGGTCTGGTCTCCTAGTCCACCATGGACGACTCGTACCCGTCTGTTACTCTTGTAGGTGTC
GTAGGTGAAGAGACCGGTACACAAGTGACACTCCTTCTTCTCCTCATGTTCTACCGGACATGTTGGAC
ATGGGACCCACAACTCTGACACCTCTACGACGGGTCTTCCGACCGTAGACCTCCACCTCACGGACT
AACCCCTCGTGGACGTACGACCGTACTCGTGGGACAAGGACACATGTGTTGTTACCGTCTGGGGGA
CCCGTACCGGAGACCGGTGTAGTCCCTGAAGGTCTAGTGACGGAGACCGGTATACCGGTACCCGGGGG
40 TTCGACCGGTCCGACGTGATGAGACCGTCTGAGTTACGGACCTCGTGGTTCCTCGGGAAGTCGACCTAGT
TCCACCTGGACGACCGGGGTACTAGTAGGTACCGTAGTTCTGGGTCCCCGGTCCGTCTTCAAGTCTGTC
GGACATGTAGTCCGTCAAGTAGTAGTACATGTGCGACCTACCGTCTTACCGTCTGGATGTCCCCGTG
TCGTGACCGTGGGACTACCACAAGAAACCGTTACACCTGTGAGACCGTAGTTCTGTTGTAGAAGTTGG
GGGGGTAGTAACGGTCTATGTAGTCCGACGTGGGGTGGGTGATGTCGTAGTCTCTGTTGGGACTCCTACCT
45 CGACTACCCGACACTGGACTTGTGACGTCGTACGGGGACCGTACCTCTCGTTCCGGTAGAGACTACGG
GTCTAGTGACGGTCTGATGAAGTGGTTGTACAAACGGTGGACCTCGGGGTCTTCCGGTCCGACGTGG
ACGTCCCGTCTCTGTTACGGACCTCCGGGGTCCAGTTGTTGGGGTTCCTACCGACGTCCACCTGAAGGT
CTTCTGGTACTTCCACTGACCCACTGGTGGGTCCCCACTTCTCGGACGACTGGTCTACATACACTTC
CTCAAGGACTAGTCTGTCGTCGGTCTACCGGTGGTACCTGGGACAAGAAGGTCTTACCGTCCACTTCC
50 ACAAGGTCCCGTGGTCTGTGCAAGTGGGGACACCACTTGTGCGACCTGGGGGGGACGACTGGTCTAT
GGACTCCTAAGTGGGGTCTCGACCCACGTGGTCTAACGGGACTCCTACCTCCACGACCCGACACTCCGG
GTCTGGACATGACT

55 SEQ ID NO: 21 Exemplified FVIII polypeptide (N6)

Length: 1670; Molecule Type: AA; Features Location/Qualifiers: SOURCE, 1..1670; MOL_TYPE, protein; ORGANISM, Homo sapiens

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSDYMQSDLGELPVDARFPFPRVPKSFPPNTSVVYKKTFLFV
 EFTDHLFNIAPRPPWMGLLGPTIQAEVYDVTVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREK
 5 EDDKVFPFGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGLAKEKTQTLHK
 FILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVMYWHVIGMTTPE
 VHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLLMDLGQFLFFCHISSHQHDGMEAYVKVDSCPEEPQLR
 MKNNEEAEDYDDDLTDSEMDVVRFDNDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLPDDRYSY
 KSQYLNNGPQIRIGRKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDITLLIIFKNQASRPYNIYPH
 10 GITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSVFVNMERDLASGLIG
 PLLICYKESVDQGRNQIMSDKRNVLFSVFDENRSWYL TENIQRFLPNPAGVQLEDPEFQASNIMHSINGY
 VFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMYEDTTLTLPFSGETVFMSENPLWILG
 CHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIP
 ENDIEKTDPPFAHRTMPKIQNVSSDMLLRQSPTPHGLSLSDLQEAKYETFSDDPSPGAIDSNNSLSE
 15 MTHFRPQLHHSMDVMFTPESGQLRLNEKLGTAAATELKKLDFKVSSTSNLISTIPSDNLAAGTDNTSSL
 GPPSMPVHYDSQLDITLFGKKSSPLTESGGPLSLSEENND SKLLESGLMNSQESSWGKNVSSREITRTTLQ
 SDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSV
 PQFKKVVFEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQ
 GAEPRKNFVKPNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAGHGR
 20 QVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRI
 RWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMS
 TLFLVYSNKCQTPLGMASGHIRDQITASGQYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIH
 GIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHP
 THYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVN
 25 NPKEWLQVDFQKTMKVTGVTTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVN
 SLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY

SEQ ID NO: 22 Exemplified FVIII polypeptide (V3)

Length: 1474; Molecule Type: AA; Features Location/Qualifiers: SOURCE, 1..1474; MOL_TYPE, protein; ORGANISM, Homo sapiens

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSDYMQSDLGELPVDARFPFPRVPKSFPPNTSVVYKKTFLF
 VEFTDHLFNIAPRPPWMGLLGPTIQAEVYDVTVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQR
 35 EKEDDKVFPFGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGLAKEKTQT
 LHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVMYWHVIGMG
 TTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLLMDLGQFLFFCHISSHQHDGMEAYVKVDSCPE
 EPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLA
 PDDRSYKSQYLNNGPQIRIGRKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDITLLIIFKNQASR
 PYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSVFVNMER
 40 DLASGLIGPLLCYKESVDQGRNQIMSDKRNVLFSVFDENRSWYL TENIQRFLPNPAGVQLEDPEFQAS
 NIMHSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMYEDTTLTLPFSGETVFM
 MENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTIP
 NSNTSNDNSVSPVLRKHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHY
 FIAAVERLWDYGMSSSPHVLNRNAQSGSVQFKKVVFEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVE
 45 DNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPKRNFKPNETKTYFWKVQHMAPTKDEFDCKAWAYF
 SDVDLEKDVHSGLIGPLLVCHTNTLNPAGHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQME
 DPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNL
 YPGVFETVEMLPKAGIWRVECLIGEHLHAGMS TLFLVYSNKCQTPLGMASGHIRDQITASGQYQWAP
 KLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGN
 50 STGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHP THYSIRSTLRMELMGCDLNSCSMPLGMESKAISDA
 QITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLTSMYVKE
 FLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEA

QDLY

SEQ ID NO: 23 Exemplified WPRE component (mWPRE)

5 Length: 600; Molecule Type: DNA; Features Location/Qualifiers: source,
1..600; mol_type, unassigned DNA; organism, Woodchuck hepatitis virus

1 GGGCCCAATC AACCTCTGGA TTACAAAATT TGTGAAAGAT TGA CTGGTAT TCTTAACTAT
61 GTTGCTCCTT TTACGCTATG TGGATACGCT GCTTTAATGC CTTTGTATCA TGCTATTGCT
121 TCCCGTATGG CTTTCATTTT CTCCTCCTTG TATAAATCCT GGTGCTGTC TCTTTATGAG
10 181 GAGTTGTGGC CCGTTGTCAG GCAACGTGGC GTGGTGTGCA CTGTGTTTGC TGACGCAACC
241 CCCACTGGTT GGGGCATTGC CACCACCTGT CAGCTCCTTT CCGGGACTTT CGCTTTCCCC
301 CTCCTTATTG CCACGGCGGA ACTCATCGCC GCCTGCCTTG CCCGCTGTG GACAGGGGCT
361 CGGCTGTTGG GCACTGACAA TTCCGTGGTG TTGTCGGGGA AATCATCGTC CTTTCCTTGG
421 CTGCTCGCCT GTGTTGCCAC CTGGATTCTG CGCGGGACGT CCTTCTGCTA CGTCCCTTCG
15 481 GCCCTCAATC CAGCGGACCT TCCTCCCCG GCCTGCTGC CGGCTCTGCG GCCTCTCCG
541 CGTCTTCGCC TTCGCCCTCA GACGAGTCGG ATCTCCCTTT GGGCCGCCCTC CCCGCAAGCT

SEQ ID NO: 24 F/HN-SIV-hCEF-soA1AT plasmid as defined in Figure 3 (pDNA1 pGM407)

20 Length: 7349; Molecule Type: DNA; Features Location/Qualifiers: source,
1..7349; mol_type, other DNA; note, pGM407; organism, synthetic construct

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCA TTGGT TATATAGCAT AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT AGTAATCAAT
25 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTTG ACGTCAATTA TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
361 AACTGCCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC GGGACTTTCC
30 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGGGTGTACG GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC CTTTGAGCC
35 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT CCTTGGCTTA
841 GAAAGCTAAT AAATTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCTT CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTTCTCTC TCTGACCCAG GCGAGAGAAA
961 TCACAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA CAGCTGAGAA
1021 GGCGTCCGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC TTGGTGAGTA
40 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG AGGCCGTAGC
1141 CGTAACCTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG CAATGGGGGC GGCTACCTCA
1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC GACTTCGCCC GAACGGAAAG
1261 AAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG AGATGGAGCG CTTGCGGCTC
1321 CATGAGAGGT TGTTGGAGAC AGAGGAGGGG TGAAAAAGAA TCATAGAAGT CCTTACCCC
45 1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA ATCTTGTGTG CGTGTATAT
1441 TGCTTGACCA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG CAGTAGCAAC AGTAAGACAA
1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA CATCTAGTGG AAAAAAGAAA
1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA ATTTTCCAGC GCAACAACAA
1621 GGAAATGCCT GGGTACATGT ACCCTTGTCA CCGCGCACCT TAAATGCGTG GGTAAAGCA
50 1681 GTAGAGGAGA AAAAATTTGG AGCAGAAATA GTACCCATTT TTTTGTTC AAGCCCTATCG
1741 AATTCCCGTT TGTGCTAGGG TTCTTAGGCT TCTTGGGGGC TGCTGGAAC TCAATGGGAG
1801 CAGCGGCGAC AGCCCTGACG GTCCAGTCTC AGCATTGTCT TGCTGGGATA CTCGACGAGC
1861 AGAAGAATCT GCTGGCGGCT GTGGAGGCTC AACAGCAGAT GTTGAAGCTG ACCATTTGGG
1921 GTGTTAAAAA CCTCAATGCC CGCGTCACAG CCCTTGAGAA GTACCTAGAG GATCAGGCAC
55 1981 GACTAAACTC CTGGGGGTGC GCATGGAAAC AAGTATGTCA TACCACAGTG GAGTGGCCCT
2041 GGACAAATCG GACTCCGGAT TGGCAAAATA TGACTTGGTT GGAGTGGGAA AGACAAATAG

	2101	CTGATTTGGA	AAGCAACATT	ACGAGACAAT	TAGTGAAGGC	TAGAGAACAA	GAGGAAAAGA
	2161	ATCTAGATGC	CTATCAGAAG	TTAACTAGTT	GGTCAGATTT	CTGGTCTTGG	TTCGATTTC
	2221	CAAAATGGCT	TAACATTTTA	AAAAATGGGAT	TTTAGTAAT	AGTAGGAATA	ATAGGGTTAA
5	2281	GATTACTTTA	CACAGTATAT	GGATGTATAG	TGAGGGTTAG	GCAGGGATAT	GTTCTCTAT
	2341	CTCCACAGAT	CCATATCCGC	GGCAATTTTA	AAAGAAAGGG	AGGAATAGGG	GGACAGACTT
	2401	CAGCAGAGAG	ACTAATTAAT	ATAATAACAA	CACAATTAGA	AATACAACAT	TTACAAACCA
	2461	AAATTCAAAA	AATTTTAAAT	TTTAGAGCCG	CGGAGATCTG	TTACATAACT	TATGGTAAAT
	2521	GGCCTGCCTG	GCTGACTGCC	CAATGACCCC	TGCCCAATGA	TGTCAATAAT	GATGTATGTT
10	2581	CCCATGTAAT	GCCAATAGGG	ACTTCCATT	GATGTCAATG	GGTGGAGTAT	TTATGGTAAC
	2641	TGCCCCACTT	GCAGTACATC	AAGTGTATCA	TATGCCAAGT	ATGCCCCCTA	TTGATGTCAA
	2701	TGATGGTAAA	TGGCCTGCCT	GGCATTATGC	CCAGTACATG	ACCTTATGGG	ACTTTCCTAC
	2761	TTGGCAGTAC	ATCTATGTAT	TAGTCATTGC	TATTACCATG	GGAAATTCAT	AGTGGAGAAG
	2821	AGCATGCTTG	AGGGCTGAGT	GCCCCTCAGT	GGGCAGAGAG	CACATGGCCC	ACAGTCCCTG
	2881	AGAAGTTGGG	GGGAGGGGTG	GGCAATTGAA	CTGGTGCCTA	GAGAAGGTGG	GGCTTGGGTA
15	2941	AACTGGGAAA	GTGATGTGGT	GTACTGGCTC	CACCTTTTTT	CCCAGGGTGG	GGGAGAACCA
	3001	TATATAAGTG	CAGTAGTCTC	TGTGAACATT	CAAGCTTCTG	CCTTCTCCCT	CCTGTGAGTT
	3061	TGCTAGCCAC	CATGCCCAGC	TCTGTGTCCT	GGGGCATTCT	GCTGCTGGCT	GGCCTGTGCT
	3121	GTCTGGTGCC	TGTGTCCCTG	GCTGAGGACC	CTCAGGGGGA	TGCTGCCCAG	AAAACAGACA
	3181	CCTCCCACCA	TGACCAGGAC	CACCCACCT	TCAACAAGAT	CACCCCCAAC	CTGGCAGAGT
20	3241	TTGCCTTCAG	CCTGTACAGA	CAGCTGGCCC	ACCAGAGCAA	CAGCACCAAC	ATCTTTTTCA
	3301	GCCCTGTGTC	CATTGCCACA	GCCTTTGCCA	TGCTGAGCCT	GGGCACCAAG	GCTGACACCC
	3361	ATGATGAGAT	CCTGGAAGGC	CTGAACTTCA	ACCTGACAGA	GATCCCTGAG	GCCCAGATCC
	3421	ATGAGGGCTT	CCAGGAACCT	CTGAGAACCC	TGAACCAGCC	AGACAGCCAG	CTGCAGTGA
	3481	CAACAGGCCA	TGGGCTGTTC	CTGTCTGAGG	GCCTGAAGCT	GGTGGACAAG	TTTCTGGAAG
25	3541	ATGTGAAGAA	GCTGTACCAC	TCTGAGGCCT	TCACAGTGAA	CTTTGGGGAC	ACAGAAGAGG
	3601	CCAAGAAACA	GATCAATGAC	TATGTGGAAG	AGGGCACCCA	GGGCAAGATT	GTGGACCTTG
	3661	TGAAAGAGCT	GGACAGGGAC	ACTGTGTTTG	CCCTTGTGAA	CTACATCTTC	TTCAAGGGCA
	3721	AGTGGGAGAG	GGCCTTTGAA	GTGAAGGACA	CTGAGGAAGA	GGACTTCCAT	GTCGACCAAG
	3781	TGACCACAGT	GAAGGTGCCA	ATGATGAAGA	GACTGGGGAT	GTTCAATATC	CAGCACTGCA
30	3841	AGAAACTGAG	CAGCTGGGTG	CTGCTGATGA	AGTACCTGGG	CAATGCTACA	GCCATATTCT
	3901	TTCTGCCTGA	TGAGGGCAAG	CTGCAGCACC	TGGAAAATGA	GCTGACCCAT	GACATCATCA
	3961	CCAAATTTCT	GGAAAATGAG	GACAGAAGAT	CTGCCAGCCT	GCATCTGCCC	AAGCTGAGCA
	4021	TCACAGGCAC	ATATGACCTG	AAGTCTGTGC	TGGGACAGCT	GGGAATCACC	AAGGTGTTCA
	4081	GCAATTGGGC	AGACCTGAGT	GGAGTGACAG	AGGAAGCCCC	TCTGAAGCTG	TCCAAGGCTG
35	4141	TGCACAAGGC	AGTGCTGACC	ATTGATGAGA	AGGGCACAGA	GGCTGTGGG	GGCATGTTTC
	4201	TGGAAGCCAT	CCCCATGTCC	ATCCCCCAG	AAGTGAAGTT	CAACAAGCCC	TTTGTGTTCC
	4261	TGATGATTGA	GCAGAACACC	AAGAGCCCC	TGTTTATGGG	CAAGGTTGTG	AACCCACCCC
	4321	AGAAATGAGG	GCCCAATCAA	CCTCTGGATT	ACAAAATTTG	TGAAAGATTG	ACTGGTATTC
	4381	TTAACTATGT	TGCTCCTTTT	ACGCTATGTG	GATACGCTGC	TTAATGCCT	TTGTATCATG
40	4441	CTATTGCTTC	CCGTATGGCT	TTCAATTTCT	CCTCCTTGTA	TAAATCCTGG	TTGCTGTCTC
	4501	TTTATGAGGA	GTTGTGGCCC	GTTGTGAGGC	AACGTGGCGT	GGTGTGCACT	GTGTTGTCTG
	4561	ACGCAACCCC	CACTGGTTGG	GGCATTGCCA	CCACCTGTCA	GCTCCTTTCC	GGGACTTTCC
	4621	CTTTCCCCCT	CCCTATTGCC	ACGGCGGAAC	TCATCGCCGC	CTGCCTTGCC	CGCTGTGGA
	4681	CAGGGGCTCG	GCTGTTGGGC	ACTGACAATT	CCGTGGTGTT	GTCGGGGAAA	TCATCGTCC
45	4741	TTCTTGGCT	GCTCGCCTGT	GTTGCCACCT	GGATTCTGCG	CGGGACGTCC	TTCTGTACG
	4801	TCCCTTCGGC	CCTCAATCCA	GCGGACCTTC	CTTCCGCGG	CCTGTGCGG	GCTCTGCGG
	4861	CTCTTCCGCG	TCTTCGCCTT	CGCCCTCAGA	CGAGTCGGAT	CTCCCTTTGG	GCCGCTCCC
	4921	CGCAAGCTTC	GCACTTTTTA	AAAGAAAAGG	GAGGACTGGA	TGGGATTTAT	TACTCCGATA
	4981	GGACGCTGGC	TTGTAACCTA	GTCTCTTACT	AGGAGACCAG	CTTGAGCCTG	GGTGTTCGCT
50	5041	GGTTAGCCTA	ACCTGGTTGG	CCACCAGGGG	TAAGGACTCC	TTGGCTTAGA	AAGCTAATAA
	5101	ACTTGCCTGC	ATTAGAGCTC	TTACGCGTCC	CGGGCTCGAG	ATCCGCATCT	CAATTAGTCA
	5161	GCAACCATAG	TCCCGCCCTT	AACCTCGGCC	ATCCCGCCCT	TAACTGCGCC	CAGTTTCGCG
	5221	CATTCTCCGC	CCCATGGCTG	ACTAATTTTT	TTTATTTATG	CAGAGGCCGA	GGCCGCCTCG
	5281	GCCTCTGAGC	TATTCAGAA	GTAGTGAGGA	GGCTTTTTTG	GAGGCCTAGG	CTTTTGCAAA
55	5341	AAGCTAACTT	GTTTATTGCA	GCTTATAATG	GTTACAAATA	AAGCAATAGC	ATCACAAATT
	5401	TCACAAATAA	AGCATTTTTT	TCATGCAATT	CTAGTTGTGG	TTTGTCCAAA	CTCATCAATG
	5461	TATCTTATCA	TGTCTGTCCG	CTTCTCGCT	CACTGACTCG	CTGCGCTCGG	TCGTTCCGGT
	5521	GCGGCGAGCG	GTATCAGCTC	ACTCAAAGGC	GGTAATACGG	TTATCCACAG	AATCAGGGGA
	5581	TAACGCAGGA	AAGAACATGT	GAGCAAAAAG	CCAGCAAAAAG	GCCAGGAACC	GTA AAAAGGC
60	5641	CGCGTTGCTG	GCGTTTTTCC	ATAGGCTCCG	CCCCCTGAC	GAGCATCACA	AAAATCGACG
	5701	CTCAAGTCAG	AGGTGGCGAA	ACCCGACAGG	ACTATAAAGA	TACCAGGCGT	TTCCCCCTGG

5 5761 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCCTT
 5821 TCTCCCTTCG GGAAGCGTGG CGCTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTCCGGT
 5881 GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCCAGC CCGACCGCTG
 5941 CGCCTTATCC GGTAACATATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT
 6001 GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT
 6061 CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGAACA GTATTTGGTA TCTGCGCTCT
 6121 GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC
 6181 CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC
 6241 TCAAGAAGAT CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAAAG AAAACTCACG
 10 6301 TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA
 6361 AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTAGAA
 6421 AAATCATCG AGCATCAAAAT GAAACTGCAA TTTATTCATA TCAGGATTAT CAATACCATA
 6481 TTTTGA AAAA AGCCGTTTCT GTAAATGAAG AGAAAACTCA CCGAGGCAGT TCCATAGGAT
 6541 GGCAAGATCC TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC AACCTATTAA
 15 6601 TTTCCCTTCG TCAAAAATAA GGTATCAAG TGAGAAATCA CCATGAGTGA CGACTGAATC
 6661 CCGTGAGAAAT GGCAACAGCT TATGCATTTT TTTCCAGACT TGTTCAACAG GCCAGCCATT
 6721 ACGCTCGTCA TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCTGT ATTGCGCCTG
 6781 AGCGAGACGA AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA
 6841 CCGGCGCAGG AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG GATATTCTTC
 20 6901 TAATACCTGG AATGCTGTTT TTCCGGGGAT CGCAGTGGTG AGTAACCATG CATCATCAGG
 6961 AGTACGGATA AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTACTCT
 7021 GACCATCTCA TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA GAAACAACCTC
 7081 TGGCGCATCG GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC
 7141 GCGAGCCCAT TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC GCGGCCATAGA
 25 7201 GCAAGACGTT TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT TTATGTAAGC
 7261 AGACAGTTTT ATTGTTTATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT
 7321 TTGAGACACA ACAATTGGTC GACGGATCC

30 SEQ ID NO: 25 F/HN-SIV-CMV-HFVIII-V3 plasmid as defined in Figure 4A (pDNA1 pGM411)

Length: 10812; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..10812; mol_type, other DNA; note, pGM411; organism, synthetic construct

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
 35 121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT AGTAATCAAT
 181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCGCG GTTACATAAC TTACGGTAAA
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
 361 AACTGCCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
 40 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC GGGATTTCCT
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGTATG GGTTTTGGCA
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG GTGGGAGGTC
 45 721 TATATAAGCA GAGCTCGCTG GCTTGTAAT CAGTCTCTTA CTAGGAGACC AGCTTGAGCC
 781 TGGGTGTTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT CCTTGGCTTA
 841 GAAAGCTAAT AAACCTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCTT CATTGACGCC
 901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTTCTCTC TCTGACCCAG GCGAGAGAAA
 961 CTCCAGCAGT GCGGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA CAGCTGAGAA
 50 1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC TTGGTGAGTA
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGAATAGGAG AGGCCGTAGC
 1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC AATGGGGGCG GCTACCTCAG
 1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAAATAC ACTTCGCCCG AACCGAAAGA
 1261 AAAAGTACCA AATTAAACAT TTAATATGGG CAGGCAAGGA GATGGAGCGC TTCGGCTTCC
 55 1321 ATGAGAGGTT GTTGGAGACA GAGGAGGGGT GTAAAAGAAT CATAGAAGTC CTCTACCCCC
 1381 TAGAACCAAC AGGATCGGAG GGCTTAAAAA GTCTGTTCAA TCTTGTGTGC GTGCTATATT
 1441 GCTTGACAAA GGAACAGAAA GTGAAAGACA CAGAGGAAGC AGTAGCAACA GTAAGACAAC
 1501 ACTGCCATCT AGTGGAAAAA GAAAAAAGTG CAACAGAGAC ATCTAGTTGA CAAAAGAAAA
 1561 ATGACAAGGG AATAGCAGCG CCACCTGGTG GCAGTCAGAA TTTTCCAGCG CAACAACAAG

	1621	GAAATGCCTG	GGTACATGTA	CCCTTGTCAC	CGCGCACCTT	AAATGCGTGG	GTAAAAGCAG
	1681	TAGAGGAGAA	AAAATTTGGA	GCAGAAATAG	TACCCATGTT	TCAAGCCCTA	TCGAATTCCC
	1741	GTTTGTGCTA	GGGTTCTTAG	GCTTCTTGGG	GGCTGCTGGA	ACTGCAATGG	GAGCAGCGGC
5	1801	GACAGCCCTG	ACGGTCCAGT	CTCAGCATTT	GCTTGCTGGG	ATACTGCAGC	AGCAGAAGAA
	1861	TCTGCTGGCG	GCTGTGGAGG	CTCAACAGCA	GATGTTGAAG	CTGACCATT	GGGGTGTTAA
	1921	AAACCTCAAT	GCCCCGCTCA	CAGCCCTTGA	GAAGTACCTA	GAGGATCAGG	CACGACTAAA
	1981	CTCCTGGGGG	TGCGCATGGA	AACAAGTATG	TCATACCACA	GTGGAGTGGC	CCTGGACAAA
	2041	TCGGACTCCG	GATTGGCAAA	ATATGACTTG	GTTGGAGTGG	GAAAGACAAA	TAGCTGATTT
	2101	GGAAAGCAAC	ATTACGAGAC	AATTAGTGAA	GGCTAGAGAA	CAAGAGGAAA	AGAATCTAGA
10	2161	TGCCTATCAG	AAGTTAACTA	GTTGGTCAGA	TTTCTGGTCT	TGGTTCGATT	TCTCAAAATG
	2221	GCTTAACATT	TTAAAAATGG	GATTTTTAGT	AATAGTAGGA	ATAATAGGGT	TAAGATTACT
	2281	TTACACAGTA	TATGGATGTA	TAGTGAGGGT	TAGGCAGGGA	TATGTTCCCTC	TATCTCCACA
	2341	GATCCATATC	CGCGGCAATT	TTAAAAAGAAA	GGGAGGAATA	GGGGGACAGA	CTTCAGCAGA
	2401	GAGACTAATT	AATATAATAA	CAACACAATT	AGAAATACAA	CATTTACAAA	CCAAAATTCA
15	2461	AAAAATTTTA	AATTTTAGAG	CCGCGGAGAT	CTCAATATTG	GCCATTAGCC	ATATTATTCA
	2521	TTGGTTATAT	AGCATAAAATC	AATATTGGCT	ATTGGCCATT	GCATACGTTG	TATCTATATC
	2581	ATAATATGTA	CATTTATATT	GGCTCATGTC	CAATATGACC	GCCATGTTGG	CATTGATTAT
	2641	TGACATGTTA	TTAATAGTAA	TCAATTACGG	GGTCATTAGT	TCATAGCCCA	TATATGGAGT
	2701	TCCGCGTTAC	ATAACTTACG	GTAAATGGCC	CGCCTGGCTG	ACCGCCCAAC	GACCCCCGCC
20	2761	CATTGACGTC	AATAATGACG	TATGTTCCCA	TAGTAACGCC	AATAGGGACT	TTCCATTGAC
	2821	GTCAATGGGT	GGAGTATTTA	CGGTAAACTG	CCCCTTGGC	AGTACATCAA	GTGTATCATA
	2881	TGCCAAGTCC	GCCCCCTATT	GACGTCAATG	ACGGTAAATG	GCCCCCTGG	CATTATGCCC
	2941	AGTACATGAC	CTTACGGGAC	TTTCTACTTT	GGCAGTACAT	CTACGTATTA	GTATCCGCTA
	3001	TTACCATGGT	GATGCGGTTT	TGGCAGTACA	CCAATGGGCG	TGGATAGCCG	TTTGACTCAC
25	3061	GGGGATTTCC	AAGTCTCCAC	CCCATTGACG	TCAATGGGAG	TTTGTTTTGG	CACCAAAATC
	3121	AACGGGACTT	TCCAAAATGT	CGTAATAACC	CCGCCCCGTT	GACGCAAATG	GGCGGTAGGC
	3181	GTGTACGGTG	GGAGGTCTAT	ATAAGCAGAG	CTCGTTTAGT	GAACCGTCAG	ATCACTAGAA
	3241	GCTTTATTGC	GGTAGTTTAT	CACAGTTAA	TTGCTAACGC	AGTCAGTGCT	TCTGACACAA
	3301	CAGTCTCGAA	CTTAAGCTGC	AGAAGTTGGT	CGTGAGGCAC	TGGGCAGGCT	AGCCCAAT
30	3361	GCAGATTGAG	CTGAGCACCT	GCTTCTTCCT	GTGCCTGCTG	AGGTTCTGCT	TCTCTGCCAC
	3421	CAGGAGATAC	TACCTGGGGG	CTGTGGAGCT	GAGCTGGGAC	TACATGCAGT	CTGACCTGGG
	3481	GGAGCTGCCT	GTGGATGCCA	GGTTCCCCC	CAGAGTGCCC	AAGAGCTTCC	CCTTCAACAC
	3541	CTCTGTGGTG	TACAAGAAGA	CCCTGTTTGT	GGAGTTCACT	GACCACCTGT	TCAACATTGC
	3601	CAAGCCCAGG	CCCCCTGGA	TGGGCTGTCT	GGGCCCCACC	ATCCAGGCTG	AGGTGTATGA
35	3661	CACTGTGGTG	ATCACCTTGA	AGAACATGGC	CAGCCACCCT	GTGAGCCTGC	ATGCTGTGGG
	3721	GGTGAGCTAC	TGGAAGGCCT	CTGAGGGGGC	TGAGTATGAT	GACCAGACCA	GCCAGAGGGA
	3781	GAAGGAGGAT	GACAAGGTGT	TCCCTGGGGG	CAGCCACACC	TATGTGTGGC	AGGTGCTGAA
	3841	GGAGAATGGC	CCCATGGCCT	CTGACCCCTT	GTGCCTGACC	TACAGCTACC	TGAGCCATGT
	3901	GGACCTGGTG	AAGGACCTGA	ACTCTGGCCT	GATTGGGGCC	CTGCTGGTGT	GCAGGGAGGG
40	3961	CAGCCTGGCC	AAGGAGAAGA	CCCAGACCCT	GCACAAGTTC	ATCCTGCTGT	TTGCTGTGTT
	4021	TGATGAGGGC	AAGAGCTGGC	ACTCTGAAAC	CAAGAACAGC	CTGATGCAGG	ACAGGGATGC
	4081	TGCCTCTGCC	AGGGCCTGGC	CCAAGATGCA	CACTGTGAAT	GGCTATGTGA	ACAGGAGCCT
	4141	GCCTGGCCTG	ATTGGCTGCC	ACAGGAAGTC	TGTGTACTGG	CATGTGATTG	GCATGGGCAC
	4201	CACCCCTGAG	GTGCACAGCA	TCTTCCTGGA	GGGCCACACC	TTCCTGGTCA	GGAAACCACAG
45	4261	GCAGGCCAGC	CTGGAGATCA	GCCCCATCAC	CTTCCTGACT	GCCCAGACCC	TGCTGATGGA
	4321	CCTGGGCCAG	TTCTGTCTGT	TCTGCCACAT	CAGCAGCCAC	CAGCATGATG	GCATGGAGGC
	4381	CTATGTGAAG	GTGGACAGCT	GCCCTGAGGA	GCCCCAGCTG	AGGATGAAGA	ACAATGAGGA
	4441	GGCTGAGGAC	TATGATGATG	ACCTGACTGA	CTCTGAGATG	GATGTGGTGA	GGTTTGATGA
	4501	TGACAACAGC	CCCAGCTTCA	TCCAGATCAG	GTCTGTGGCC	AAGAAGCACC	CCAAGACCTG
50	4561	GGTGCACTAC	ATTGCTGCTG	AGGAGGAGGA	CTGGGACTAT	GCCCCCTGG	TGCTGGCCCC
	4621	TGATGACAGG	AGCTACAAGA	GCCAGTACCT	GAACAATGGC	CCCCAGAGGA	TTGGCAGGAA
	4681	GTACAAGAAG	GTCAAGTTCA	TGGCCTACAC	TGATGAAACC	TTCAAGACCA	GGGAGGCCAT
	4741	CCAGCATGAG	TCTGGCATCC	TGGGCCCCCT	GCTGTATGGG	GAGGTGGGGG	ACACCTGTCT
	4801	GATCATCTTC	AAGAACCAGG	CCAGCAGGCC	CTACAACATC	TACCCCCATG	GCATCACTGA
55	4861	TGTGAGGCCC	CTGTACAGCA	GGAGGCTGCC	CAAGGGGGTG	AAGCACCTGA	AGGACTTCCC
	4921	CATCCTGCCT	GGGGAGATCT	TCAAGTACAA	GTGGACTGTG	ACTGTGGAGG	ATGGCCCCAC
	4981	CAAGTCTGAC	CCCAGGTGCG	TGACCAGATA	CTACAGCAGC	TTTGTGAACA	TGGAGAGGGA
	5041	CCTGGCCTCT	GGCCTGATTG	GCCCCCTGCT	GATCTGCTAC	AAGGAGTCTG	TGGACCAGAG
	5101	GGGCAACCAG	ATCATGTCTG	ACAAGAGGAA	TGTGATCCTG	TTCTCTGTGT	TTGATGAGAA
60	5161	CAGGAGCTGG	TACCTGACTG	AGAACATCCA	GAGGTTCTCT	CCCAACCCTG	CTGGGGTGCA
	5221	GCTGGAGGAC	CCTGAGTTCC	AGGCCAGCAA	CATCATGCAC	AGCATCAATG	GCTATGTGTT

	5281	TGACAGCCTG	CAGCTGTCTG	TGTGCCTGCA	TGAGGTGGCC	TACTGGTACA	TCCTGAGCAT
	5341	TGGGGCCCAG	ACTGACTTCC	TGTCGTGTGT	CTTCTCTGGC	TACACCTTCA	AGCACAAAGAT
	5401	GGTGTATGAG	GACACCTGA	CCCTGTTCCC	CTTCTCTGGG	GAGACTGTGT	TCATGAGCAT
5	5461	GGAGAACCCT	GGCCTGTGGA	TTCTGGGCTG	CCACAACCTCT	GACTTCAGGA	ACAGGGGCAT
	5521	GACTGCCCTG	CTGAAAGTCT	CCAGCTGTGA	CAAGAACACT	GGGGACTACT	ATGAGGACAG
	5581	CTATGAGGAC	ATCTCTGCCT	ACCTGCTGAG	CAAGAACAAT	GCCATTGAGC	CCAGGAGCTT
	5641	CAGCCAGAAT	GCCACTAATG	TGTCTAACAA	CAGCAACACC	AGCAATGACA	GCAATGTGTC
	5701	TCCCCCAGTG	CTGAAGAGGC	ACCAGAGGGA	GATCACCAGG	ACCACCCTGC	AGTCTGACCA
10	5761	GGAGGAGATT	GACTATGATG	ACACCATCTC	TGTGGAGATG	AAGAAGGAGG	ACTTTGACAT
	5821	CTACGACGAG	GACGAGAACC	AGAGCCCCAG	GAGCTTCCAG	AAGAAGACCA	GGCACTACTT
	5881	CATTGCTGCT	GTGGAGAGGC	TGTGGGACTA	TGGCATGAGC	AGCAGCCCCC	ATGTGCTGAG
	5941	GAACAGGGCC	CAGTCTGGCT	CTGTGCCCCA	GTCAAGAAG	GTGGTGTTC	AGGAGTTTAC
	6001	TGATGGCAGC	TTCACCCAGC	CCCTGTACAG	AGGGGAGCTG	AATGAGCACC	TGGGCCCTGCT
	6061	GGGGCCCTAC	ATCAGGGCTG	AGGTGGAGGA	CAACATCATG	GTGACCTTCA	GGAAACCAGGC
15	6121	CAGCAGGCCC	TACAGCTTCT	ACAGCAGCCT	GATCAGCTAT	GAGGAGGACC	AGAGGCAGGG
	6181	GGCTGAGCCC	AGGAAGAACT	TTGTGAAGCC	CAATGAAACC	AAGACCTACT	TCTGGAAGGT
	6241	GCAGCACCAC	ATGGCCCCCA	CCAAGGATGA	GTTTGACTGC	AAGGCCTGGG	CCTACTTCTC
	6301	TGATGTGGAC	CTGGAGAAGG	ATGTGCATCT	TGGCCTGATT	GGCCCCCTGC	TGGTGTGCCA
	6361	CACCAACACC	CTGAACCCCTG	CCCATGGCAG	GCAGGTGACT	GTGCAGGAGT	TTGCCCTGTT
20	6421	CTTCACCATC	TTTGATGAAA	CCAAGAGCTG	GTACTTCACT	GAGAACATGG	AGAGGAACTG
	6481	CAGGGCCCCC	TGCAACATCC	AGATGGAGGA	CCCCACCTTC	AAGGAGAACT	ACAGGTTCCA
	6541	TGCCATCAAT	GGCTACATCA	TGGACACCCT	GCCTGGCCTG	GTGATGGCCC	AGGACCAGAG
	6601	GATCAGGTGG	TACCTGCTGA	GCATGGGCAG	CAATGAGAAC	ATCCACAGCA	TCCACTTCTC
	6661	TGGCCATGTG	TTCATGTGTA	GGAAGAAGGA	GGAGTACAAG	ATGGCCCTGT	ACAACCTGTA
25	6721	CCCTGGGGTG	TTTGAGACTG	TGGAGATGCT	GCCCAGCAAG	GCTGGCATCT	GGAGGGTGGA
	6781	GTGCCTGATT	GGGGAGCACC	TGCATGCTGG	CATGAGCACC	CTGTTCCCTGG	TGTACAGCAA
	6841	CAAGTGCCAG	ACCCCCCTGG	GCATGGCCTC	TGGCCACATC	AGGGACTTCC	AGATCACTGC
	6901	CTCTGGCCAG	TATGGCCAGT	GGGCCCCCAA	GCTGGCCAGG	CTGCACTACT	CTGGCAGCAT
	6961	CAATGCCTGG	AGCACCAAGG	AGCCCTTCAG	CTGGATCAAG	GTGGACCTGC	TGGTCCCCAT
30	7021	GATCATCCAT	GGCATCAAGA	CCCAGGGGGC	CAGGCAGAAG	TTCAGCAGCC	TGTACATCAG
	7081	CCAGTTTCATC	ATCATGTACA	GCCTGGATGG	CAAGAAGTGG	CAGACCTACA	GGGGCAACAG
	7141	CACTGGCACC	CTGATGTTGT	TCTTTGGCAA	TGTGGACAGC	TCTGGCATCA	AGCACAAACAT
	7201	CTTCAACCCC	CCCATCATTG	CCAGATACAT	CAGGCTGCAC	CCCACCCACT	ACAGCATCAG
	7261	GAGCACCCCTG	AGGATGGAGC	TGATGGGCTG	TGACCTGAAC	AGCTGCAGCA	TGCCCTTGGG
35	7321	CATGGAGAGC	AAGGCCATCT	CTGATGCCCA	GATCACTGCC	AGCAGTACT	TGCCAACAT
	7381	GTTTGCCACC	TGGAGCCCCA	GCAAGGCCAG	GCTGCACCTG	CAGGGCAGGA	GCAATGCCTG
	7441	GAGGCCCCAG	GTCAACAACC	CCAAGGAGTG	GCTGCAGGTG	GACTTCCAGA	AGACCATGAA
	7501	GGTGAAGTGG	GTGACCACCC	AGGGGGTGAA	GAGCCTGCTG	ACCAGCATGT	ATGTGAAGGA
	7561	GTTCTTGATC	AGCAGCAGCC	AGGATGGCCA	CCAGTGGACC	CTGTTCTTCC	AGAATGGCAA
40	7621	GGTGAAGGTG	TTCCAGGGCA	ACCAGGACAG	CTTCACCCCT	GTGGTGAACA	GCCTGGACCC
	7681	CCCCCTGCTG	ACCAGATACC	TGAGGATTCA	CCCCCAGAGC	TGGGTGCACC	AGATTGCCCT
	7741	GAGGATGGAG	GTGCTGGGCT	GTGAGGCCCA	GGACCTGTAC	TGAGCGGCCG	CGGGCCCAAT
	7801	CAACCTCTGG	ATTACAAAAT	TTGTGAAAGA	TTGACTGGTA	TTCTTAACTA	TGTTGCTCCT
	7861	TTTACGCTAT	GTGGATACGC	TGCTTTAATG	CCTTTGTATC	ATGCTATTGC	TTCCCGTATG
45	7921	GCTTTTCATTT	TCTCCTCCTT	GTATAAATCC	TGGTTGCTGT	CTCTTTATGA	GGAGTTGTGG
	7981	CCCGTTGTCA	GGCAACGTGG	CGTGGTGTGC	ACTGTGTTTG	CTGACGCAAC	CCCCACTGGT
	8041	TGGGGCATTG	CCACCACCTG	TCAGCTCCTT	TCCGGGACTT	TCGCTTTCCC	CCTCCCTATT
	8101	GCCACGGCGG	AACTCATCGC	CGCCTGCCTT	GCCCCTGCTG	GGACAGGGGC	TCGGCTGTTG
	8161	GGCACTGACA	ATTCCGTGGT	GTTGTCGGGG	AAATCATCGT	CCTTTCCCTG	GCTGCTCGCC
50	8221	TGTGTTGCCA	CCTGGATTCT	GCGCGGGACG	TCCTTCTGCT	ACGTCCCTTC	GGCCCTCAAT
	8281	CCAGCGGACC	TTCTTTCCCG	CGGCCTGCTG	CCGGCTCTGC	GGCCTCTTCC	CGCTCTTCGC
	8341	CTTCGCCCTC	AGACGAGTCG	GATCTCCCTT	TGGGCCGCCCT	CCCCGCAAGC	TTCCGCACTT
	8401	TTAAAAGAAA	AGGGAGGACT	GGATGGGATT	TATTACTCCG	ATAGGACGCT	GGCTTGTAAC
	8461	TCAGTCTCTT	ACTAGGAGAC	CAGCTTGAGC	CTGGGTGTTT	GCTGGTTAGC	CTAACCTGGT
55	8521	TGGCCACCAG	GGGTAAGGAC	TCCTTGGCTT	AGAAAGCTAA	TAAACTTGCC	TGCATTAGAG
	8581	CTCTTACGCG	TCCCGGGCTC	GAGATCCGCA	TCTCAATTAG	TCAGCAACCA	TAGTCCCGCC
	8641	CCTAACTCCG	CCCATCCCGC	CCCTAACTCC	GCCCAGTTCC	GCCCATTCTC	CGCCCATGGG
	8701	CTGACTAATT	TTTTTTAATT	ATGCAGAGGC	CGAGGCCGCC	TCGGCCTCTG	AGCTATTCCA
	8761	GAAGTAGTGA	GGAGGCTTTT	TTGGAGGCCT	AGGCTTTTGC	AAAAAGCTAA	CTTGTTTATT
60	8821	GCAGCTTATA	ATGTTTACAA	ATAAAGCAAT	AGCATCACAA	ATTTACAAAA	TAAAGCATT
	8881	TTTTCACTGC	ATTCTAGTTG	TGGTTTGTCC	AAACTCATCA	ATGTATCTTA	TCATGTCTGT

8941 CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTCGTTTCG GCTGCGGCGA GCGGTATCAG
 9001 CTCCTCAAAA GGCGGTAATA CGGTATATCCA CAGAATCAGG GGATAACGCA GGAAAGAACA
 9061 TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG CTGGCGTTTT
 5 9121 TCCATAGGCT CCGCCCCCTT GACGAGCATC AAAAAAATCG ACGCTCAAGT CAGAGGTGGC
 9181 GAAACCCGAC AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC CTCGTGCGCT
 9241 CTCCTGTTCC GACCCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT TCGGGAAGCG
 9301 TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC GTTCGCTCCA
 9361 AGCTGGGCTG TGTGCACGAA CCCCCGTTT AGCCCGACCG CTGCGCCTTA TCCGGTAACT
 10 9421 ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA GCCACTGGTA
 9481 ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG TGGTGGCCTA
 9541 ACTACGGCTA CACTAGAAGA ACAGTATTTG GTATCTGCGC TCTGCTGAAG CCAGTTACCT
 9601 TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT AGCGGTGGTT
 9661 TTTTGTGTTT CAAGCAGCAG ATTACGCGCA GAAAAAAGG ATCTCAAGAA GATCCTTTGA
 9721 TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAACTC ACGTTAAGGG ATTTTGGTCA
 15 9781 TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA TTAAAAATGA AGTTTTAAAT
 9841 CAATCTAAAG TATATATGAG TAACTTGGT CTGACAGTTA GAAAACTCA TCGAGCATCA
 9901 AATGAACTG CAATTTATTC ATATCAGGAT TATCAATACC ATATTTTGA AAAAGCCGTT
 9961 TCTGTAATGA AGGAGAAAAAC TCACCGAGGC AGTTCCATAG GATGGCAAGA TCCTGGTATC
 10021 GGTCTGCGAT TCCGACTCGT CCAACATCAA TACAACCTAT TAATTTCCCC TCGTCAAAAA
 20 10081 TAAGGTTATC AAGTGAGAAA TCACCATGAG TGACGACTGA ATCCGGTGAG AATGGCAACA
 10141 GCTTATGCAT TTCTTTCCAG ACTTGTTCAA CAGGCCAGCC ATTACGCTCG TCATCAAAAT
 10201 CACTCGCATC AACCAAACCG TTATTCATTC GTGATTGCGC CTGAGCGAGA CGAAATACGC
 10261 GATCGCTGTT AAAAGGACAA TTACAAACAG GAATCGAATG CAACCGGCGC AGGAACACTG
 10321 CCAGCGCATC AACAATATTT TCACCTGAAT CAGGATATTC TTCTAATACC TGAATGCTG
 25 10381 TTTTCCGGG GATCGCAGTG GTGAGTAACC ATGCATCATC AGGAGTACGG ATAAATGCT
 10441 TGATGGTTCG AAGAGGCATA AATTCGTCA GCCAGTTTAG TCTGACCATC TCATCTGTAA
 10501 CATCATGGC AACGCTACCT TTGCCATGTT TCAGAAACAA CTCTGGCGCA TCGGGCTTCC
 10561 CATACAATCG ATAGATTGTC GCACCTGATT GCCCGACATT ATCGCGAGCC CATTTATACC
 10621 CATATAATC AGCATCCATG TTGGAATTTA ATCGCGGCC TATTCCTGTT TTTCCGTT
 30 10681 GAATATGGCT CATAACACCC CTTGTATTAC TGTATTATGTA AGCAGACAGT TTTATTGTTT
 10741 ATGATGATAT ATTTTATCT TGTGCAATGT AACATCAGAG ATTTTGAGAC ACAACAATTG
 10801 GTCGACGGAT CC

35 SEQ ID NO: 26 F/HN-SIV-hCEF-HFVIII-V3 plasmid as defined in Figure 4B (pDNA1 pGM413)

Length: 10519; Molecule Type: DNA; Features Location/Qualifiers: source,
 1..10519; mol_type, other DNA; note, pGM413; organism, synthetic construct
 1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT
 61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
 40 121 ATGTCCAATA TGACCGCCAT GTTGGCATAT ATTATTGACT AGTTATTAAT AGTAATCAAT
 181 TACGGGGTCA TTAGTTTATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
 241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT
 301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
 361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
 45 421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC GGGACTTTCC
 481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
 541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT
 601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA
 661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG GTGGGAGGTC
 50 721 TATATAAGCA GAGCTCGCTG GCTTGTAATT CAGTCTCTTA CTAGGAGACC AGCTTGAGCC
 781 TGGGTGTTTC CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT CCTTGGCTTA
 841 GAAAGCTAAT AAACCTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCTT CATTGACGCC
 901 TCACTCTCTT GAACGGGAAT CTTCTTACT GGGTTCTCTC TCTGACCCAG GCGAGAGAAA
 961 CTCCAGCAGT GGCGCCCGAA CAGGGAATTG AGTGAGAGTG TAGGCACGTA CAGCTGAGAA
 55 1021 GGCGTCCGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC TTGGTGAGTA
 1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG AGGCCGTAGC
 1141 CGTAATACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC AATGGGGGCG GCTACCTCAG
 1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAATACG ACTTCGCCCG AACGGAAAGA
 1261 AAAAGTACCA AATTAAACAT TTAATATGGG CAGGCAAGGA GATGGAGCGC TTCGGCTTCC

	1321	ATGAGAGGTT	GTTGGAGACA	GAGGAGGGGT	GTAAGAAGAT	CATAGAAGTC	CTCTACCCCC
	1381	TAGAACCAAC	AGGATCGGAG	GGCTTAAAAA	GTCTGTTCAA	TCTTGTGTGC	GTGCTATATT
	1441	GCTTGACAAA	GGAACAGAAA	GTGAAAGACA	CAGAGGAAGC	AGTAGCAACA	GTAAGACAAC
5	1501	ACTGCCATCT	AGTGGAAAAA	GAAAAAAGTG	CAACAGAGAC	ATCTAGTGGA	CAAAAGAAAA
	1561	ATGACAAGGG	AATAGCAGCG	CCACCTGGTG	GCAGTCAGAA	TTTTCCAGCG	CAACAACAAG
	1621	GAAATGCCTG	GGTACATGTA	CCCTTGTCAC	CGCGCACCTT	AAATGCGTGG	GTAAAAGCAG
	1681	TAGAGGAGAA	AAAATTTGGA	GCAGAAATAG	TACCCATGTT	TCAAGCCCTA	TGAATTTCCC
	1741	GTTTGTGCTA	GGGTTCTTAG	GCTTCTTGGG	GGCTGCTGGA	ACTGCAATGG	GAGCAGCGGC
10	1801	GACAGCCCTG	ACGGTCCAGT	CTCAGCATTT	GCTTGCTGGG	ATACTGCAGC	AGCAGAGAAA
	1861	TCTGCTGGCG	GCTGTGGAGG	CTCAACAGCA	GATGTTGAAG	CTGACCATT	GGGGTGTAA
	1921	AAACCTCAAT	GCCCGCGTCA	CAGCCCTTGA	GAAGTACCTA	GAGGATCAGG	CACGACTAAA
	1981	CTCCTGGGGG	TGCGCATGGA	AACAAGTATG	TCATACCACA	GTGGAGTGAC	CCTGGACAAA
	2041	TCGGACTCCG	GATTGGCAAA	ATATGACTTG	GTTGGAGTGG	GAAAGACAAA	TAGCTGATTT
15	2101	GGAAAGCAAC	ATTACGAGAC	AATTAGTGAA	GGCTAGAGAA	CAAGAGGAAA	AGAATCTAGA
	2161	TGCCTATCAG	AAGTTAACTA	GTTGGTCAGA	TTCTTGGTCT	TGGTTCGATT	TCTCAAATG
	2221	GCTTAACATT	TTAAAAATGG	GATTTTTAGT	AATAGTAGGA	ATAATAGGGT	TAAGATTACT
	2281	TTACACAGTA	TATGGATGTA	TAGTGAGGGT	TAGGCAGGGA	TATGTTCCCT	TATCTCCACA
	2341	GATCCATATC	CGCGGCAATT	TTAAAAGAAA	GGGAGGAATA	GGGGGACAGA	CTTCAGACAGA
20	2401	GAGACTAATT	AATATAATAA	CAACACAATT	AGAAATACAA	CATTTACAAA	CCAAAATTCA
	2461	AAAAATTTTA	AATTTTAGAG	CCGCGGAGAT	CTGTTACATA	ACTTATGGTA	AATGGCCTGC
	2521	CTGGCTGACT	GCCCAATGAC	CCCTGCCCAA	TGATGTCAAT	AATGATGTAT	GTTCCCATGT
	2581	AATGCCAATA	GGGACTTTCC	ATTGATGTCA	ATGGGTGGAG	TATTTATGGT	AACTGCCAC
	2641	TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	AGTATGCCCC	CTATTGATGT	CAATGATGGT
25	2701	AAATGGCCTG	CCTGGCATT	TGCCAGTAC	ATGACCTTAT	GGGACTTTCC	TACTTGGCAG
	2761	TACATCTATG	TATTAGTCAT	TGCTATTACC	ATGGGAATTC	ACTAGTGGAG	AAGAGCATGC
	2821	TTGAGGGCTG	AGTGCCCCCTC	AGTGGGCAGA	GAGCACATGG	CCCACAGTCC	CTGAGAAAGT
	2881	GGGGGGAGGG	GTGGGCAATT	GAACCTGGTG	CTAGAGAAGG	TGGGGCTTGG	GTAAACTGGG
	2941	AAAGTGATGT	GGTGTACTGG	CTCCACCTTT	TTCCCCAGGG	TGGGGGAGAA	CCATATATAA
30	3001	GTGCAGTAGT	CTCTGTGAAC	ATTCAAGCTT	CTGCCCTTCTC	CCTCCTGTGA	GTTTGTGAGC
	3061	CACCAATGCA	GATTGAGCTG	AGCACCTGCT	TCTTCTGTG	CCTGCTGAGG	TTCTGCTTCT
	3121	CTGCCACCAG	GAGATACTAC	CTGGGGGCTG	TGGAGCTGAG	CTGGGACTAC	ATGCAGTCTG
	3181	ACCTGGGGGA	GCTGCCGTGTG	GATGCCAGGT	TCCCCCCCAG	AGTGCCCAAG	AGCTTCCCTC
	3241	TCAACACCTC	TGTGGTGTAC	AAGAAGACCC	TGTTTGTGGA	GTTCACCTGAC	CACCTGTTCA
35	3301	ACATTGCCAA	GCCCAGGCC	CCCTGGATGG	GCCTGCTGGG	CCCCACCATC	CAGGCTGAGG
	3361	TGTATGACAC	TGTGGTGATC	ACCCTGAAGA	ACATGGCCAG	CCACCCTGTG	AGCCTGTATG
	3421	CTGTGGGGGT	GAGCTACTGG	AAGGCTCTG	AGGGGGCTGA	GTATGATGAC	CAGACCAGCC
	3481	AGAGGGAGAA	GGAGGATGAC	AAGGTGTTCC	CTGGGGGCAG	CCACACCTAT	GTGTGGCAGG
	3541	TGCTGAAGGA	GAATGGCCCC	ATGGCCTCTG	ACCCCTGTG	CCTGACCTAC	AGCTACCTGA
40	3601	GCCATGTGGA	CCTGGTGAAG	GACCTGAAC	CTGGCCTGAT	TGGGGCCCTG	CTGGTGTGCA
	3661	GGGAGGGCAG	CCTGGCCAAG	GAGAAGACCC	AGACCCTGCA	CAAGTTCATC	CTGCTGTTTG
	3721	CTGTGTTTGA	TGAGGGCAAG	AGCTGGCACT	CTGAAACCAA	GAACAGCCTG	ATGCAGGACA
	3781	GGGATGCTGC	CTCTGCCAGG	GCCTGGCCCA	AGATGCACAC	TGTGAATGGC	TATGTGAACA
	3841	GGAGCCTGCC	TGGCCTGATT	GGCTGCCACA	GGAAGTCTGT	GTACTGGCAT	GTGATTGGCA
45	3901	TGGGCACCAC	CCCTGAGGTG	CACAGCATCT	TCCTGGAGGG	CCACACCTTC	CTGGTGAGGA
	3961	ACCACAGGCA	GGCCAGCCTG	GAGATCAGCC	CCATCACCTT	CCTGACTGCC	CAGACCTTGC
	4021	TGATGGACCT	GGGCCAGTTC	CTGCTGTTCT	GCCACATCAG	CAGCCACCAG	CATGATGGCA
	4081	TGGAGGCCCTA	TGTGAAGGTG	GACAGCTGCC	CTGAGGAGCC	CCAGCTGAGG	ATGAAGAACA
	4141	ATGAGGAGGC	TGAGGACTAT	GATGATGACC	TGACTGACTC	TGAGATGGAT	GTGGTGAGGT
50	4201	TTGATGATGA	CAACAGCCCC	AGCTTCATCC	AGATCAGGTC	TGTGGCCAAG	AAGCACCCCA
	4261	AGACCTGGGT	GCACTACATT	GCTGCTGAGG	AGGAGGACTG	GGACTATGCC	CCCCTGGTGC
	4321	TGGCCCTTGA	TGACAGGAGC	TACAAGAGCC	AGTACCTGAA	CAATGGCCCC	CAGAGGATTG
	4381	CGAGGAAGTA	CAAGAAGGTC	AGGTTTCACTG	CCTACACTGA	TGAAACCTTC	AAGACCAGGG
	4441	AGGCCATCCA	GCATGAGTCT	GGCATCCTGG	GCCCCCTGCT	GTATGGGGAG	GTGGGGGACA
55	4501	CCCTGCTGAT	CATCTTCAAG	AACCAGGCCA	GCAGGCCCTA	CAACATCTAC	CCCCATGGCA
	4561	TCATGATGAT	GAGGCCCTG	TACAGCAGGA	GGCTGCCCAA	GGGGGTGAAG	CACCTGAAGG
	4621	ACTTCCCCAT	CCTGCCTGGG	GAGATCTTCA	AGTACAAAGT	GACTGTGACT	GTGGAGGATG
	4681	CCCCACCAA	GTCTGACCCC	AGGTGCCTGA	CCAGATACTA	CAGCAGCTTT	GTGAACATGG
	4741	AGAGGGACCT	GGCCTCTGGC	CTGATTGGCC	CCCTGCTGAT	CTGCTACAAG	GAGCTGTGG
60	4801	ACCAGAGGGG	CAACCAGATC	ATGTCTGACA	AGAGGAATGT	GATCCTGTTC	TCTGTGTTTG
	4861	ATGAGAACAG	GAGCTGGTAC	CTGACTGAGA	ACATCCAGAG	GTTCTTCCCC	AACCTGCTG
	4921	GGGTGCAGCT	GGAGGACCCT	GAGTTCCAGG	CCAGCAACAT	CATGCACAGC	ATCAATGGCT

	4981	ATGTGTTTGA	CAGCCTGCAG	CTGTCTGTGT	GCCTGCATGA	GGTGGCCTAC	TGGTACATCC
	5041	TGAGCATTGG	GGCCCAGACT	GACTTCCTGT	CTGTGTTCTT	CTCTGGCTAC	ACCTTCAAGC
	5101	ACAAGATGGT	GTATGAGGAC	ACCGTGACCC	TGTTCCCTTT	CTCTGGGGAG	ACTGTGTTCA
5	5161	TGAGCATGGA	GAACCCTGGC	CTGTGGATTG	TGGGCTGCCA	CAACTCTGAC	TTCAGGAACA
	5221	GGGGCATGAC	TGCCCTGCTG	AAAGTCTCCA	GCTGTGACAA	GAACACTGGG	GACTACTATG
	5281	AGGACAGCTA	TGAGGACATC	TCTGCCTACC	TGCTGAGCAA	GAACAATGCC	ATTGAGCCCA
	5341	GGAGCTTCAG	CCAGAATGCC	ACTAATGTGT	CTAACAACAG	CAACACCAGC	AATGACAGCA
	5401	ATGTGTCTCC	CCCAGTGCTG	AAGAGGCACC	AGAGGGAGAT	CACCAGGACC	ACCCTGCAGT
10	5461	CTGACCAGGA	GGAGATTGAC	TATGATGACA	CCATCTCTGT	GGAGATGAAG	AAGGAGGACT
	5521	TTGACATCTA	CGACGAGGAC	GAGAACCAGA	GCCCCAGGAG	CTTCCAGAAG	AAGACCAGGC
	5581	ACTACTTTCAT	TGCTGCTGTG	GAGAGGCTGT	GGGACTATGG	CATGAGCAGC	AGCCCCCATG
	5641	TGCTGAGGAA	CAGGGCCAG	TCTGGCTCTG	TGCCCCAGTT	CAAGAAGCTG	GTGTTCCAGG
	5701	AGTTCACTGA	TGGCAGCTTC	ACCCAGCCCC	TGTACAGAGG	GGAGCTGAAT	GAGCACCTGG
15	5761	GCCTGCTGGG	CCCCTACATC	AGGGCTGAGG	TGGAGGACAA	CATCATGGTG	ACCTTCAGGA
	5821	ACCAGGCCAG	CAGGCCCTAC	AGCTTCTACA	GCAGCCTGAT	CAGCTATGAG	GAGGACCAGA
	5881	GGCAGGGGGC	TGAGCCCAGG	AAGAACCTTG	TGAAGCCCCA	TGAAACCAAG	ACCTACTTCT
	5941	GGAGGTTGCA	GCACCACATG	GCCCCACCA	AGGATGAGTT	TGACTGCAAG	GCCTGGGCCT
	6001	ACTTCTCTGA	TGTGGACCTG	GAGAAGGATG	TGCACTCTGG	CCTGATTGGC	CCCTTGCTGG
20	6061	TGTGCCACAC	CAACACCCTG	AACCCTGCCC	ATGGCAGGCA	GGTGA CTGTG	CAGGAGTTTG
	6121	CCCTGTTCTT	CACCATCTTT	GATGAAACCA	AGAGCTGGTA	CTTCACTGAG	AACATGGAGA
	6181	GGAACTGCAG	GGCCCCCTGC	AACATCCAGA	TGGAGGACCC	CACCTTCAAG	GAGA ACTTACA
	6241	GGTTCCATGC	CATCAATGGC	TACATCATGG	ACACCCTGCC	TGGCCTGGTG	ATGGCCCAGG
	6301	ACCAGAGGAT	CAGGTGGTAC	CTGCTGAGCA	TGGGCAGCAA	TGAGAACATC	CACAGCATCC
	6361	ACTTCTCTGG	CCATGTGTTC	ACTGTGAGGA	AGAAGGAGGA	GTACAAGATG	GCCCTGTACA
25	6421	ACCTGTACCC	TGGGGTGT TT	GAGACTGTGG	AGATGCTGCC	CAGCAAGGCT	GGCATCTGGA
	6481	GGGTGGAGTG	CCTGATTGGG	GAGCACCTGC	ATGCTGGCAT	GAGCACCTTG	TTCCTGGTGT
	6541	ACAGCAACAA	GTGCCAGACC	CCCTTGGGCA	TGGCCTCTGG	CCACATCAGG	GACTTCCAGA
	6601	TCACTGCCCT	TGGCCAGTAT	GGCCAGTGGG	CCCCCAAGCT	GGCCAGGCTG	CACTACTCTG
30	6661	GCAGCATCAA	TGCCTGGAGC	ACCAAGGAGC	CCTTCAGCTG	GATCAAGTGT	GACCTGTGGG
	6721	CCCCCATGAT	CATCCATGGC	ATCAAGACCC	AGGGGGCCAG	GCAGAAGTTC	AGCAGCCTGT
	6781	ACATCAGCCA	GTTTCATCAT	ATGTACAGCC	TGGATGGCAA	GAAGTGGCAG	ACCTACAGGG
	6841	GCAACAGCAC	TGGCACCTTG	ATGGTGT TCT	TTGGCAATGT	GGACAGCTCT	GGCATCAAGC
	6901	ACAACATCTT	CAACCCCCCT	ATCATTTGCCA	GATACATCAG	GCTGCACCCC	ACCCACTACA
35	6961	GCATCAGGAG	CACCCTGAGG	ATGGAGCTGA	TGGGCTGTGA	CCTGAACAGC	TGCAGCATGC
	7021	CCCTGGGCA	GGAGAGCAAG	GCCATCTCTG	ATGCCAGATG	CACTGCCAGC	AGCTACTTCA
	7081	CCAACATGTT	TGCCACCTGG	AGCCCCAGCA	AGGCCAGGCT	GCACCTGCAG	GGCAGGAGCA
	7141	ATGCCTGGAG	GCCCCAGGTC	AACAACCCCA	AGGAGTGGCT	GCAGGTGGAC	TTCCAGAAGA
	7201	CCATGAAGGT	GACTGGGGTG	ACCACCCAGG	GGGTGAAGAG	CCTGCTGACC	AGCATGTATG
40	7261	TGAAGGAGTT	CCTGATCAGC	AGCAGCCAGG	ATGGCCACCA	GTGGACCCTG	TTCTTCCAGA
	7321	ATGGCAAGGT	GAAGGTGTTT	CAGGGCAACC	AGGACAGCTT	CACCCCTGTG	GTGAACAGCC
	7381	TGGACCCCCC	CCTGCTGACC	AGATACCTGA	GGATTACCCC	CCAGAGCTTG	GTGCACCAGA
	7441	TTGCCCTGAG	GATGGAGGTG	CTGGGCTGTG	AGGCCAGGGA	CCTGTACTGA	GCGGCCGCGG
	7501	GCCCAATCAA	CCTCTGGATT	ACAAAATTTG	TGAAAGATTG	ACTGGTATTC	TTAACTATGT
45	7561	TGCTCCTTTT	ACGCTATGTG	GATACGCTGC	TTTAATGCCT	TTGTATCATG	CTATTGCTTC
	7621	CCGTATGGCT	TTCAATTTCT	CCTCCTTGTA	TAAATCCTGG	TTGCTGTCTC	TTTATGAGGA
	7681	GTTGTGGCCC	GTTGTGAGGC	AACGTGGCGT	GGTGTGCACT	GTGTTTGCTG	ACGCAACCCC
	7741	CACGTGGTTG	GGCATTGCCA	CCACCTGTCA	GCTCCTTTCC	GGGACTTTTC	CTTTCCCTTC
	7801	CCCTATTGCC	ACGGCGGAAC	TCATCGCCGC	CTGCCTTGCC	CGCTGCTGGA	CAGGGGCTCG
50	7861	GCTGTTGGGC	ACTGACAAAT	CCGTGGTGTT	GTGGGGGAAA	TCATCGTCCT	TTCCTTGGCT
	7921	GCTCGCCTGT	GTTGCCACCT	GGATTCTGCG	CGGGACGTCC	TTCTGCTACG	TCCCTTGGGC
	7981	CCTCAATCCA	GCGGACCTTC	CTTCCC GCGG	CCTGCTGCGG	GCTCTGCGGC	CTCTTCCGCG
	8041	TGTTGCGCTT	CGCCCTCAGA	CGAGTCGGAT	CTCCCTTTGG	GCCGCTTCCC	CGCAAGCTTC
	8101	GCAC TTTT TA	AAAGAAAAGG	GAGGACTGGA	TGGGATTTAT	TACTCCGATA	GGACGCTGGC
	8161	TTGTA ACTCA	GTCTCTTACT	AGGAGACCAG	CTTGAGCCTG	GGTGTTGCTG	GGTTAGCCTA
55	8221	ACCTGGTTGG	CCACCAGGGG	TAAGGACTCC	TTGGCTTAGA	AAGCTAATAA	ACTTGCCTGC
	8281	ATTAGAGCTC	TTACGCGTCC	CGGGCTCGAG	ATCCGCATCT	CAATTAGTCA	GCAACCATAG
	8341	TCCCGCCCCC	AACTCCGCCC	ATCCCGCCCC	TAAC TCCGCC	CAGTTCCGCC	CATTCTCCGC
	8401	CCCATGGCTG	ACTAATTTT	TTTATTTATG	CAGAGGCCGA	GGCCGCTCCG	GCCCTTGAGC
	8461	TATTCCAGAA	GTAGTGAGGA	GGCTTTTTTG	GAGGCCTAGG	CTTTTGCAAA	AAGCTAACTT
60	8521	GTTTATTGCA	GCTTATAATG	GTTACAAATA	AAGCAATAGC	ATCACAAATT	TCACAAATAA
	8581	AGCATTTTTT	TCACTGCATT	CTAGTTGTGG	TTTGTCCAAA	CTCATCAATG	TATCTTATCA

	8641	TGTCTGTCCG	CTTCCTCGCT	CACTGACTCG	CTGCGCTCGG	TCGTTTCGGCT	GCGGCGAGCG
	8701	GTATCAGCTC	ACTCAAAGGC	GGTAATACGG	TTATCCACAG	AATCAGGGGA	TAACGCAGGA
	8761	AAGAACATGT	GAGCAAAAGG	CCAGCAAAAG	GCCAGGAACC	GTAAAAAGGC	CGCGTTGCTG
	8821	GCGTTTTTCC	ATAGGCTCCG	CCCCCTGAC	GAGCATCACA	AAAATCGACG	CTCAAGTCAG
5	8881	AGGTGGCGAA	ACCCGACAGG	ACTATAAAGA	TACCAGGCGT	TTCCCCCTGG	AAGCTCCCTC
	8941	GTGCGCTCTC	CTGTTCCGAC	CCTGCCGCTT	ACCGGATACC	TGTCCGCTT	TCTCCCTTCG
	9001	GGAAGCGTGG	CGCTTTCTCA	TAGCTCACGC	TGTAGGTATC	TCAGTTCGGT	GTAGGTCTGT
	9061	CGCTCCAAGC	TGGGCTGTGT	GCACGAACCC	CCCGTTCAGC	CCGACCGCTG	CGCCTTATCC
	9121	GGTAACATATC	GTCTTGAGTC	CAACCCGGTA	AGACACGACT	TATCGCCACT	GGCAGCAGCC
10	9181	ACTGGTAACA	GGATTAGCAG	AGCGAGGTAT	GTAGGCGGTG	CTACAGAGTT	CTTGAAGTGG
	9241	TGGCCTAACT	ACGGCTACAC	TAGAAGAACA	GTATTTGGTA	TCTGCGCTCT	GCTGAAGCCA
	9301	GTTACCTTCG	GAAAAAGAGT	TGGTAGCTCT	TGATCCGGCA	AACAAACCAC	CGCTGGTAGC
	9361	GGTGGTTTTT	TTGTTTGCAA	GCAGCAGATT	ACGCGCAGAA	AAAAAGGATC	TCAAGAAGAT
	9421	CCTTTGATCT	TTTCTACGGG	GTCTGACGCT	CAGTGGAAACG	AAAACCTCAG	TTAAGGGATT
15	9481	TTGGTCATGA	GATTATCAAA	AAGGATCTTC	ACCTAGATCC	TTTTAAATTA	AAAATGAAGT
	9541	TTTAAATCAA	TCTAAAGTAT	ATATGAGTAA	ACTTGGTCTG	ACAGTTAGAA	AAACTCATCG
	9601	AGCATCAAA	GAAACTGCAA	TTTATTCATA	TCAGGATTAT	CAATACCATA	TTTTTGAAAA
	9661	AGCCGTTTTCT	GTAATGAAGG	AGAAAACTCA	CCGAGGCAGT	TCCATAGGAT	GGCAAGATCC
	9721	TGGTATCGGT	CTGCGATTCC	GACTCGTCCA	ACATCAATAC	AACCTATTAA	TTTCCCCCTCG
20	9781	TCAAAAATAA	GGTTATCAAG	TGAGAAATCA	CCATGAGTGA	CGACTGAATC	CGGTGAGAAT
	9841	GGCAACAGCT	TATGCATTTC	TTTCCAGACT	TGTTCAACAG	GCCAGCCATT	ACGCTCGTCA
	9901	TCAAAATCAC	TCGCATCAAC	CAAACCGTTA	TTCATTCTGT	ATTGCGCCTG	AGCGAGACGA
	9961	AATACGCGAT	CGCTGTAA	AGGACAATTA	CAAACAGGAA	TCCAATGCAA	CCGGCGCAGG
25	10021	AACACTGCCA	GCGCATCAAC	AATATTTTCA	CCTGAATCAG	GATATTCTTC	TAATACCTGG
	10081	AATGCTGTTT	TTCCGGGGAT	CGCAGTGGTG	AGTAACCATG	CATCATCAGG	AGTACGGATA
	10141	AAATGCTTGA	TGGTCGGAAG	AGGCATAAAT	TCCGTCAGCC	AGTTTAGTCT	GACCATCTCA
	10201	TCTGTAACAT	CATTGGCAAC	GCTACCTTTG	CCATGTTTCA	GAAACAACCTC	TGGCGCATCG
	10261	GGCTTCCCAT	ACAATCGATA	GATTGTGCGA	CCTGATTGCC	CGACATTATC	GCGAGCCCAT
	10321	TTATACCCAT	ATAAATCAGC	ATCCATGTTG	GAATTTAATC	GCGGCCTAGA	GCAAGACGTT
30	10381	TCCCGTTGAA	TATGGCTCAT	AACACCCCTT	GTATTACTGT	TTATGTAAGC	AGACAGTTTT
	10441	ATTGTTTCATG	ATGATATATT	TTTATCTTGT	GCAATGTAAC	ATCAGAGATT	TTGAGACACA
	10501	ACAATTGGTC	GACGGATCC				

35 SEQ ID NO: 27 F/HN-SIV-CMV-HFVIII-N6-co plasmid as defined in Figure 4C (pDNA1 pGM412)

Length: 11400; Molecule Type: DNA; Features Location/Qualifiers: source,		1..11400; mol_type, other DNA; note, pGM412; organism, synthetic construct					
	1	GGTACCTCAA	TATTGGCCAT	TAGCCATATT	ATTCATTGGT	TATATAGCAT	AAATCAATAT
	61	TGGCTATTGG	CCATTGCATA	CGTTGTATCT	ATATCATAAT	ATGTACATTT	ATATTGGCTC
40	121	ATGTCCAATA	TGACCGCCAT	GTTGGCATTG	ATTATTGACT	AGTTATTAAT	AGTAATCAAT
	181	TACGGGGTCA	TTAGTTCATA	GCCCATATAT	GGAGTTCCGC	GTTACATAAC	TTACGGTAAA
	241	TGGCCCGCCT	GGCTGACCGC	CCAACGACCC	CCGCCCATTG	ACGTCAATAA	TGACGTATGT
	301	TCCCATAGTA	ACGCCAATAG	GGACTTTCCA	TTGACGTCAA	TGGGTGGAGT	ATTTACGGTA
	361	AAC TGCCAC	TTGGCAGTAC	ATCAAGTGTA	TCATATGCCA	AGTCCGCCCC	CTATTGACGT
45	421	CAATGACGGT	AAATGGCCCG	CCTGGCATT	TGCCCAGTAC	ATGACCTTAC	GGGACTTTCC
	481	TACTTGGCAG	TACATCTACG	TATTAGTCAT	CGCTATTACC	ATGGTGATGC	GGTTTTGGCA
	541	GTACACCAAT	GGGCGTGGAT	AGCGGTTTGA	CTCACGGGGA	TTTCCAAGTC	TCCACCCCAT
	601	TGACGTCAAT	GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	GACTTTCCAA	ATGTGCGTAA
	661	CAACTGCGAT	CGCCCGCCCT	TTTGACGCCA	ATGGGCGGTA	GGCGTGTACG	GATGGAGGTC
50	721	TATATAAGCA	GAGCTCGCTG	GCTTGTAAC	CAGTCTCTTA	CTAGGAGACC	AGCTTGAGCC
	781	TGGGTGTTTCG	CTGGTTAGCC	TAACCTGGTT	GGCCACCAGG	GGTAAGGACT	CCTTGGCTTA
	841	GAAAGCTAAT	AAACTTGCCT	GCATTAGAGC	TTATCTGAGT	CAAGTGTCTT	CATTGACGCC
	901	TCACTCTCTT	GAACGGGAAT	CTTCCTTACT	GGTTCTCTC	TCTGACCCAG	GCGAGAGAAA
	961	CTCCAGCAGT	GGCGCCCGAA	CAGGGACTTG	AGTGAGAGTG	TAGGCACGTA	CAGCTGAGAA
55	1021	GGCGTCGGAC	GCGAAGGAAG	CGCGGGGTGC	GACGCGACCA	AGAAGGAGTA	TTGGTGAGTA
	1081	GGCTTCTCGA	GTGCCGGGAA	AAAGCTCGAG	CCTAGTTAGA	GGACTAGGAG	AGGCCGTAGC
	1141	CGTAACTACT	CTGGGCAAGT	AGGGCAGGCG	GTGGGTACGC	AATGGGGGCG	GCTACCTCAG
	1201	CACTAAATAG	GAGACAATTA	GACCAATTTG	AGAAAATACG	ACTTCGCCCG	AACGGAAAGA
	1261	AAAAGTACCA	AATTAAACAT	TTAATATGGG	CAGGCAAGGA	GATGGAGCGC	TTCCGGCTCC

	1321	ATGAGAGGTT	GTTGGAGACA	GAGGAGGGGT	GTAAGAAGAT	CATAGAAGTC	CTCTACCCCC
	1381	TAGAACCAAC	AGGATCGGAG	GGCTTAAAAA	GTCTGTTCAA	TCTTGTGTGC	GTGCTATATT
	1441	GCTTGCACAA	GGAACAGAAA	GTGAAAGACA	CAGAGGAAGC	AGTAGCAACA	GTAAGACAAC
5	1501	ACTGCCATCT	AGTGGAAAAA	GAAAAAAGTG	CAACAGAGAC	ATCTAGTGGA	CAAAAGAAAA
	1561	ATGACAAGGG	AATAGCAGCG	CCACCTGGTG	GCAGTCAGAA	TTTTCCAGCG	CAACAACAAG
	1621	GAAATGCCTG	GGTACATGTA	CCCTTGTCAC	CGCGCACCTT	AAATGCGTGG	GTAAAAAGCAG
	1681	TAGAGGAGAA	AAAATTTGGA	GCAGAAATAG	TACCCATGTT	TCAAGCCCTA	TCGAATTCCC
	1741	GTTTGTGCTA	GGGTTCTTAG	GCTTCTTGGG	GGCTGCTGGA	ACTGCAATGG	GAGCAGCGGC
10	1801	GACAGCCCTG	ACGGTCCAGT	CTCAGCATTT	GCTTGCTGGG	ATACTGCAGC	AGCAGAGAAA
	1861	TCTGCTGGCG	GCTGTGGAGG	CTCAACAGCA	GATGTTGAAG	CTGACCATT	GGGGTGTAA
	1921	AAACCTCAAT	GCCCGCGTCA	CAGCCCTTGA	GAAGTACCTA	GAGGATCAGG	CACGACTAAA
	1981	CTCCTGGGGG	TGCGCATGGA	AACAAGTATG	TCATACCACA	GTGGAGTGGC	CCTGGACAAA
	2041	TCGGACTCCG	GATTGGCAAA	ATATGACTTG	GTTGGAGTGG	GAAAGACAAA	TAGCTGATTT
15	2101	GGAAAGCAAC	ATTACGAGAC	AATTAGTGAA	GGCTAGAGAA	CAAGAGGAAA	AGAATCTAGA
	2161	TGCCTATCAG	AAGTTAACTA	GTTGGTCAGA	TTTCTGGTCT	TGGTTCGATT	TCTCAAATATG
	2221	GCTTAACATT	TTAAAAATGG	GATTTTTAGT	AATAGTAGGA	ATAATAGGGT	TAAGATTACT
	2281	TTACACAGTA	TATGGATGTA	TAGTGAGGGT	TAGGCAGGGA	TATGTTCCCT	TATCTCCACA
	2341	TGCCATATC	CGCGGCAATT	TTAAAAGAAA	GGGAGGAATA	GGGGGACATC	CTTCAGACAGA
	2401	GAGACTAATT	AATATAATAA	CAACACAATT	AGAAATACAA	CATTTACAAA	CCAAAATTCA
20	2461	AAAAATTTTA	AATTTTAGAG	CCGCGGAGAT	CTCAATATTG	GCCATTAGCC	ATATTATTCA
	2521	TTGGTTTATAT	AGCATAAATC	AATATTGGCT	ATTGGCCATT	GCATACGTTG	TATCTATATC
	2581	ATAATATGTA	CATTTATATT	GGCTCATGTC	CAATATGACC	GCCATGTTGG	CATTGATTAT
	2641	TGACTAGTTA	TTAATAGTAA	TCAATTACGG	GGTCATTAGT	TCATAGCCCA	TATATGGAGT
25	2701	TCCGCGTTAC	ATAACTTACG	GTAAATGGCC	CGCCTGGCTG	ACCGCCCAAC	GACCCCGGCC
	2761	CATTGACGTC	AATAATGACG	TATGTTCCCA	TAGTAACGCC	AATAGGGACT	TTCCATTGAC
	2821	GTCAATGGGT	GGAGTATTTA	CGGTAAACTG	CCCACTTGGC	AGTACATCAA	GTGTATCATA
	2881	TGCCAAGTCC	GCCCCCTATT	GACGTCAATG	ACGGTAAATG	GCCCCGCTGG	CATTATGCCC
	2941	AGTACATGAC	CTTACGGGAC	TTTCCTACTT	GGCAGTACAT	CTACGTATTA	GTCTATCGCTA
30	3001	TTACCATTGGT	GATGCGGTTT	TGGCAGTACA	CCAATGGGCG	TGGATAGCCG	TTTGACTCAC
	3061	GGGGATTTCC	AAGTCTCCAC	CCCATTGACG	TCAATGGGAG	TTTGTTTTGG	CACCAAAATC
	3121	AACGGGACTT	TCCAAAATGT	CGTAATAACC	CCGCCCCGTT	GACGCAAATG	GGCGGTAGGC
	3181	GTGTACGGTG	GGAGGTCTAT	ATAAGCAGAG	CTCGTTTAGT	GAACCGTCAG	ATCATTAGAA
	3241	GCTTTATTGC	GGTAGTTTAT	CACAGTTAAA	TTGCTAACGC	AGTCAGTGCT	TCTGACACAA
35	3301	CAGTCTCGAA	CTTAAGCTGC	AGAAGTTGGT	CGTGAGGCAC	TGGGCAGGCT	AGCCACCAAT
	3361	GCAGATTGAG	CTGAGCACCT	GCTTCTTCC	GTGCTGCTG	AGGTTCTGCT	TCTCTGCCAC
	3421	CAGGAGATAC	TACCTGGGGG	CTGTGGAGCT	GAGCTGGGAC	TACATGCAGT	CTGACCTGGG
	3481	GGAGCTGCCT	GTGGATGCCA	GGTTCCCCC	CAGAGTGCCC	AAGAGCTTCC	CCTTCAACAC
	3541	CTCTGTGGTG	TACAAGAAGA	CCCTGTTTGT	GGAGTTCACT	GACCACCTGT	TCAACATTGC
40	3601	CAAGCCCAGG	CCCCCTGGA	TGGGCTGCT	GGGCCCCACC	ATCCAGGCTG	AGGTGTATGA
	3661	CACGTGTGGT	ATCACCTGTA	AGAACATGGC	CAGCCACCC	GTGAGCCTGC	ATGCTGTGGG
	3721	GGTGAGCTAC	TGGAAGCCCT	CTGAGGGGGC	TGAGTATGAT	GACCAGACCA	GCCAGAGGGA
	3781	GAAGGAGGAT	GACAAGGTGT	TCCCTGGGGG	CAGCCACACC	TATGTGTGGC	AGGTGCTGAA
	3841	GGAGAATGGC	CCCATGGCCT	CTGACCCCT	GTGCCTGACC	TACAGCTACC	TGAGCCATGT
45	3901	GGACCTGGTG	AAGGACCTGA	ACTCTGGCCT	GATTGGGGCC	CTGCTGGTGT	GCAGGGAGGG
	3961	CAGCCTGGCC	AAGGAGAAGA	CCCAGACCCT	GCACAAGTTC	ATCCTGCTGT	TTGCTGTGTT
	4021	TGATGAGGGC	AAGAGCTGGC	ACTCTGAAAC	CAAGAACAGC	CTGATGCAGG	ACAGGGATGC
	4081	TGCCTCTGCC	AGGGCCTGGC	CCAAGATGCA	CACGTGTGAAT	GGCTATGTGA	ACAGGAGCCT
	4141	GCCTGGCCTG	ATTGGCTGCC	ACAGGAAGTC	TGTGTACTGG	CATGTGATTG	GCATGGGCAC
50	4201	CACCCCTGAG	GTGCACAGCA	TCTTCTTGGA	GGGCCACACC	TTCCTGGTCA	GGAACCACAG
	4261	GCAGGCCAGC	CTGGAGATCA	GCCCCATCAC	CTTCTGACT	GCCCAGACCC	TGCTGATGGA
	4321	CCTGGGCCAG	TTCTGCTGT	TCTGCCACAT	CAGCAGCCAC	CAGCATGATG	GCATGGAGGC
	4381	CTATGTGAAG	GTGGACAGCT	GCCCTGAGGA	GCCCCAGCTG	AGGATGAAGA	ACAAATGAGGA
	4441	GGCTGAGGAC	TATGATGATG	ACCTGACTGA	CTCTGAGATG	GATGTGGTGA	GGTTTGTATGA
55	4501	TGACAACAGC	CCCAGCTTCA	TCCAGATCAG	GTCTGTGGCC	AAGAAGCACC	CCAAGACCTG
	4561	GGTGCACTAC	ATTGCTGCTG	AGGAGGAGGA	CTGGGACTAT	GCCCCCTGG	TGCTGGCCCC
	4621	TGATGACAGG	AGCTACAAGA	GCCAGTACCT	GAACAATGGC	CCCCAGAGGA	TTGGCAGGAA
	4681	GTACAAGAAG	GTCAGGTTCA	TGGCCTACAC	TGATGAAACC	TTCAAGACCA	GGGAGGCCAT
	4741	CCAGCATGAG	TCTGGCATCC	TGGGCCCCCT	GCTGTATGGG	GAGGTGGGGG	ACACCTTGCT
60	4801	GATCATCTTC	AAGAACCAGG	CCAGCAGGCC	CTACAACATC	TACCCCCATG	GCATCACTGA
	4861	TGTGAGGCCC	CTGTACAGCA	GGAGGCTGCC	CAAGGGGGTG	AAGCACCTGA	AGGACTTCCC
	4921	CATCCTGCCT	GGGGAGATCT	TCAAGTACAA	GTGGACTGTG	ACTGTGGAGG	ATGGCCCCAC

	4981	CAAGTCTGAC	CCCAGGTGCC	TGACCAGATA	CTACAGCAGC	TTTGTGAACA	TGGAGAGGGA
	5041	CCTGGCCTCT	GGCCTGATTG	GGCCCTTGCT	GATCTGCTAC	AAGGAGTCTG	TGGACCAGAG
	5101	GGGCAACCAG	ATCATGTCTG	ACAAGAGGAA	TGTGATCCTG	TTCTCTGTGT	TTGATGAGAA
5	5161	CAGGAGCTGG	TACCTGACTG	AGAACATCCA	GAGGTTCCCTG	CCCAACCCCTG	CTGGGGTGCA
	5221	GCTGGAGGAC	CCTGAGTTCC	AGGCCAGCAA	CATCATGCAC	AGCATCAATG	GCTATGTGTT
	5281	TGACAGCCTG	CAGCTGTCTG	TGTGCCTGCA	TGAGGTGGCC	TACTGGTACA	TCCTGAGCAT
	5341	TGGGGCCCAG	ACTGACTTCC	TGTCTGTGTT	CTTCTCTGGC	TACACCTTCA	AGCACAAAGAT
	5401	GGTGTATGAG	GACACCCTGA	CCCTGTTCCC	CTTCTCTGGG	GAGACTGTGT	TCATGAGCAT
10	5461	GGAGAACCCCT	GGCCTGTGGA	TTCTGGGCTG	CCACAACCTCT	GACTTCAGGA	ACAGGGGCAT
	5521	GACTGCCCTG	CTGAAAGTCT	CCAGCTGTGA	CAAGAACACT	GGGGACTACT	ATGAGGACAG
	5581	CTATGAGGAC	ATCTCTGCCT	ACCTGCTGAG	CAAGAACAAT	GCCATTGAGC	CCAGGAGCTT
	5641	CAGCCAGAAC	AGCAGGCACC	CCAGCACCAG	GCAGAAGCAG	TTCAATGCCA	CCACCATCCC
	5701	TGAGAATGAC	ATAGAGAAGA	CAGACCCATG	GTTTGCCAC	CGGACCCCCA	TGCCCCAAGAT
	5761	CCAGAATGTG	AGCAGCTCTG	ACCTGCTGAT	GCTGCTGAGG	CAGAGCCCCA	CCCCCATGG
15	5821	CCTGAGCCTG	TCTGACCTGC	AGGAGGCCAA	GTATGAAACC	TTCTCTGATG	ACCCACGCC
	5881	TGGGGCCATT	GACAGCAACA	ACAGCCTGTC	TGAGATGACC	CACCTCAGGC	CCCAGCTGCA
	5941	CCACTCTGGG	GACATGGTGT	TCACCCCTGA	GTCTGGCCTG	CAGCTGAGGC	TGAATGAGAA
	6001	GCTGGGCACC	ACTGCTGCCA	CTGAGCTGAA	GAAGCTGGAC	TTCAAAGTCT	CCAGCACCAG
	6061	CAACAACCTG	ATCAGCACCA	TCCCTCTGTA	CAACCTGGCT	GCTGGCACTG	ACAACACCAG
20	6121	CAGCCTGGGC	CCCCCAGCA	TGCCTGTGCA	CTATGACAGC	CAGCTGGACA	CCACCCTGTT
	6181	TGGCAAGAAG	AGCAGCCCCC	TGACTGAGTC	TGGGGGCCCC	CTGAGCCTGT	CTGAGGAGAA
	6241	CAATGACAGC	AAGCTGCTGG	AGTCTGGCCT	GATGAACAGC	CAGGAGAGCA	GCTGGGGCAA
	6301	GAATGTGAGC	AGCAGGGAGA	TCACCAGGAC	CACCTGCGAG	TCTGACCAGG	AGGAGATGTA
	6361	CTATGATGAC	ACCATCTCTG	TGGAGATGAA	GAAGGAGGAC	TTTGACATCT	ACGACGAGGA
25	6421	CGAGAACCAG	AGCCCCAGGA	GCTTCCAGAA	GAAGACCAGG	CACTACTTCA	TTGCTGCTGT
	6481	GGAGAGGCTG	TGGGACTATG	GCATGAGCAG	CAGCCCCCAT	GTGCTGAGGA	ACAGGGCCCA
	6541	GTCTGGCTCT	GTGCCCCAGT	TCAAGAAGGT	GGTGTTCCAG	GAGTTCACTG	ATGGCAGCTT
	6601	CACCCAGCCC	CTGTACAGAG	GGGAGCTGAA	TGAGCACCTG	GGCCTGCTGG	GCCCCATCAT
	6661	CAGGGCTGAG	GTGGAGGACA	ACATCATGGT	GACCTTCAGG	AACCAGGCCA	GCAGGCCCTA
30	6721	CAGCTTCTAC	AGCAGCCTGA	TCAGCTATGA	GGAGGACCAG	AGGCAGGGGG	CTGAGCCCAG
	6781	GAAGAACTTT	GTGAAGCCCA	ATGAAACCAA	GACCTACTTC	TGGAAGGTGC	AGCACCACAT
	6841	GGCCCCCACC	AAGGATGAGT	TTGACTGCAA	GGCCTGGGCC	TACTTCTCTG	ATGTGGACCT
	6901	GGAGAAGGAT	GTGCACTCTG	GCCTGATTGG	CCCCCTGCTG	GTGTGCCACA	CCAACACCCCT
	6961	GAACCCTGCC	CATGGCAGGC	AGGTGACTGT	GCAGGAGTTT	GCCCTGTTCT	TCACCATCTT
35	7021	TGATGAAACC	AAGAGCTGGT	ACTTCACTGA	GAACATGGAG	AGGAACCTGA	GGGCCCCCTG
	7081	CAACATCCAG	ATGGAGGACC	CCACCTTCAA	GGAGAACTAC	AGGTTCCATG	CCATCAATGG
	7141	CTACATCATG	GACACCCTGC	CTGGCCTGGT	GATGGCCCAG	GACCAGAGGA	TCAGGTGGTA
	7201	CCTGCTGAGC	ATGGGCAGCA	ATGAGAACAT	CCACAGCATC	CACCTTCTCTG	GCCATGTGTT
	7261	CACTGTGAGG	AAGAAGGAGG	AGTACAAGAT	GGCCCTGTAC	AACCTGTACC	CTGGGGTGTT
40	7321	TGAGACTGTG	GAGATGCTGC	CCAGCAAGGC	TGGCATCTGG	AGGGTGGAGT	GCCTGATTGG
	7381	GGAGCACCTG	CATGCTGGCA	TGAGCACCCCT	GTTCCCTGGT	TACAGCAACA	AGTGCCAGAC
	7441	CCCCCTGGGC	ATGGCCTCTG	GCCACATCAG	GGACTTCCAG	ATCACTGCCT	CTGGCCAGTA
	7501	TGGCCAGTGG	GGCCCCAAGC	TGGCCAGGCT	GCACTACTCT	GGCAGCATCA	ATGCCTGGAG
	7561	CACCAAGGAG	CCCTTCAGCT	GGATCAAGGT	GGACCTGCTG	GCCCCCATGA	TCATCCATGG
45	7621	CATCAAGACC	CAGGGGGCCA	GGCAGAAGTT	CAGCAGCCTG	TACATCAGCC	AGTTCATCAT
	7681	CATGTACAGC	CTGGATGGCA	AGAAGTGGCA	GACCTACAGG	GGCAACAGCA	CTGGCACCCT
	7741	GATGGTGTTT	TTTGGCAATG	TGGACAGCTC	TGGCATCAAG	CACAACATCT	TCAACCCCCC
	7801	CATCATTGCC	AGATACATCA	GGCTGCACCC	CACCCACTAC	AGCATCAGGA	GCACCCTGAG
	7861	GATGGAGCTG	ATGGGCTGTG	ACCTGAACAG	CTGCAGCATG	CCCCTGGGCA	TGGAGAGCAA
50	7921	GGCCATCTCT	GATGCCCAGA	TCACTGCCAG	CAGCTACTTC	ACCAACATGT	TTGCCACCTG
	7981	GAGCCCCAGC	AAGGCCAGGC	TGCACCTGCA	GGGCAGGAGC	AATGCCTGGA	GGCCCCAGGT
	8041	GAACAACCCC	AAGGAGTGGC	TGCAGGTGGA	CTTCCAGAAG	ACCATGAAGG	TGACTGGGGT
	8101	GACCACCCAG	GGGGTGAAGA	GCCTGCTGAC	CAGCATGTAT	GTGAAGGAGT	TCCTGATCAG
	8161	CAGCAGCCAG	GATGGCCACC	AGTGGACCCT	GTTCTTCCAG	AATGGCAAGG	TGAAGGTGTT
55	8221	CCAGGGCAAC	CAGGACAGCT	TCACCCCTGT	GGTGAACAGC	CTGGACCCCC	CCCTGCTGAC
	8281	CAGATACCTG	AGGATTACCC	CCCAGAGCTG	GGTGACCAG	ATTGCCCTGA	GGATGGAGGT
	8341	GCTGGGCTGT	GAGGCCCAGG	ACCTGTACTG	AGCGGCCGCG	GGCCCAATCA	ACCTCTGGAT
	8401	TACAAAATTT	GTGAAAAGATT	GACTGGTATT	CTTAACATATG	TGCTCCTTTT	TACGCTATGT
	8461	GGATACGCTG	CTTTAATGCC	TTTGTATCAT	GCTATTGCTT	CCCGTATGGC	TTTCATTTTC
60	8521	TCCTCCTTGT	ATAAATCCTG	GTTGCTGTCT	CTTTATGAGG	AGTTGTGGCC	CGTTGTICAGG
	8581	CAACGTGGCG	TGGTGTGCAC	TGTGTTTGCT	GACGCAACCC	CCACTGGTTG	GGGCATTGCC

```

5 8641 ACCACCTGTC AGCTCCTTTC CGGGACTTTC GCTTTCCCCC TCCCTATTGC CACGGCGGAA
8701 CTCATCGCCG CCTGCCTTGC CCGTGCTGG ACAGGGGCTC GGCTGTTGGG CACTGACAAAT
8761 TCCGTGGTGT TGTCGGGGAA ATCATCGTCC TTCCCTTGGC TGCTCGCCTG TGTGCCCACC
8821 TGGATTCTGC GCGGGACGTC CTTCTGCTAC GTCCCTTCCG CCCTCAATCC AGCGGACCTT
8881 CCTTCCCGCG GCCTGCTGCC GGCTCTGCGG CCTCTTCCGC GTCTTCGCCCT TCGCCCTCAG
8941 ACGAGTCGGA TCTCCCTTTC GGCCGCCTCC CCGCAAGCTT CGCACTTTTT AAAAGAAAAG
9001 GGAGGACTGG ATGGGATTTA TTACTCCGAT AGGACGCTGG CTTGTAATC AGTCTCTTAC
9061 TAGGAGACCA GCTTGAGCCT GGGTGTTCGC TGGTTAGCCT AACCTGGTTG GCCACCAGGG
9121 GTAAGGACTC CTTGGCTTAG AAAGCTAATA AACTTGCCTG CATTAGAGCT CTTACGCGTC
10 9181 CCGGGCTCGA GATCCGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC TAACTCCGCC
9241 CATCCCAGCC CTAATCCCG CCAGTTCGCG CCATTCTCCG CCCCATGGCT GACTAATTTT
9301 TTTTATTTAT GCAGAGGCCG AGGCCGCCTC GGCTCTGAG CTATTCAGAG AGTAGTGAGG
9361 AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA AAAGCTAACT TGTTTATTGC AGCTTATAAT
9421 GGTTACAAAT AAAGCAATAG CATCACAAT TTCACAAATA AAGCATTTTT TTCACTGCAT
15 9481 TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGTCC GCTTCTCGC
9541 TCACTGACTC GCTGCGCTCG GTCGTTCCGG TCGGCGGAGC GGTATCAGCT CACTCAAAGG
9601 CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAG
9661 GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC CATAGGCTCC
9721 GCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG
20 9781 GACTATAAAG ATACCAGGCG TTTCCCTCTG GAAGCTCCCT CGTGCCTCTT CCTGTTCCGA
9841 CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGAAGCGTGG GCGCTTTCTC
9901 ATAGCTCAGC CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG
9961 TGCACGAACC CCCCGTTCAG CCCGACGCT GCGCTTATC CGGTAACAT CGTCTTGAGT
25 10021 CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA
10081 GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTGAAGTG GTGGCCTAAC TACGGCTACA
10141 CTAGAAGAAC AGTATTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG
10201 TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA
10261 AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG
10321 GGTCTGACGC TCAGTGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA
30 10381 AAAGGATCTT CACCTAGATC CTTTAAATTT AAAAATGAAG TTTTAAATCA ATCTAAAGTA
10441 TATATGAGTA AACTTGGTCT GACAGTTAGA AAAACTCATC GAGCATCAAA TGAACCTGCA
10501 ATTTATTCAT ATCAGGATTA TCAATACCAT ATTTTGAAGA AAGCCGTTTC TGTAATGAAG
10561 GAGAAAACTC ACCGAGGCAG TTCCATAGGA TGGCAAGATC CTGGTATCGG TCTGCGATTC
10621 CGACTCGTCC AACATCAATA CAACCTATTA ATTTCCCTTC GTCAAAAATA AGGTTATCAA
35 10681 GTGAGAAATC ACCATGAGTG ACGACTGAAT CCGGTGAGAA TGGCAACAGC TTATGCATTT
10741 CTTTCCAGAC TTGTTCAACA GGCCAGCCAT TACGCTCGTC ATCAAAATCA CTCGCATCAA
10801 CCAAACCGTT ATTCAATCGT GATTGCGCCT GAGCGAGACG AAATACGCGA TCGCTGTTAA
10861 AAGGACAATT ACAACAGGA ATCGAATGCA ACCGGCGCAG GAACACTGCC AGCGCATCAA
10921 CAATATTTTC ACCTGAATCA GGATATTCTT CTAATACCTG GAATGCTGTT TTTCCGGGGA
40 10981 TCGCAGTGGT GAGTAACCAT GCATCATGAG GAGTACGGAT AAAATGCTTG ATGGTCGGAA
11041 GAGGCATAAA TTCCGTCAGC CAGTTTAGTC TGACCATCTC ATCTGTAACA TCATTGGCAA
11101 CGCTACCTTT GCCATGTTTC AGAAACAAC CTGGCGCATC GGGCTTCCCA TACAATCGAT
11161 AGATTGTGCG ACCTGATTGC CCGACATTAT CGCGAGCCCA TTTATACCCA TATAAATCAG
11221 CATCCATGTT GGAATTTAAT CGCGGCCTAG AGCAAGACGT TTCCCGTTGA ATATGGCTCA
45 11281 TAACACCCCT TGTATTACTG TTTATGTAAG CAGACAGTTT TATTGTTTAT GATGATATAT
11341 TTTTATCTTG TGCAATGTAA CATCAGAGAT TTTGAGACAC AACAAATTGGT CGACGGATCC

```

SEQ ID NO: 28 F/HN-SIV-hCEF-HFVIII-N6-co plasmid as defined in Figure 4D (pDNA1 pGM414)

```

50 Length: 11108; Molecule Type: DNA; Features Location/Qualifiers: source,
1..11108; mol_type, other DNA; note, pGM414; organism, synthetic construct
1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT AGTAATCAAT
55 181 TACGGGGTCA TTAGTTCATA GCCATATAT GGAGTTCCGC GTTACATAAC TTACGTTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTC ACGTCAATAA TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGAATTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
361 AACTGCCCCA TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTG TGCCAGTAC ATGACCTTAC GGGACTTTCC

```

	481	TACTTGGCAG	TACATCTACG	TATTAGTCAT	CGCTATTACC	ATGGTGATGC	GGTTTTGGCA
	541	GTACACCAAT	GGGCGTGGAT	AGCGGTTTGA	CTCACGGGGA	TTTCCAAGTC	TCCACCCCAT
	601	TGACGTCAAT	GGGAGTTTGT	TTTGGCACCA	AAATCAACGG	GACTTTCCAA	AATGTCTTAA
5	661	CAACTGCGAT	CGCCCCCCCC	GTTGACGCAA	ATGGGCGGTA	GGCGTGTACG	GTGGGAGGTC
	721	TATATAAGCA	GAGCTCGCTG	GCTTGTAAC	CAGTCTCTTA	CTAGGAGACC	AGCTTGAGCC
	781	TGGGTGTTTC	CTGGTTAGCC	TAACCTGGTT	GGCCACCAGG	GGTAAGGACT	CCTTGGCTTA
	841	GAAAGCTAAT	AAACTTGCCT	GCATTAGAGC	TTATCTGAGT	CAAGTGTCTT	CATTGACGCC
	901	TCACTCTCTT	GAACGGGAAT	CTTCCTTACT	GGGTTCTCTC	TCTGACCCAG	GCGAGAGAAA
10	961	CTCCAGCAGT	GGCGCCCGAA	CAGGGAAGTT	AGTGAGAGTG	TAGGCACGTA	CAGCTGAGAA
	1021	GGCGTCCGAC	GCGAAGGAAG	CGCGGGGTGC	GACGCGACCA	AGAAGGAGAC	TTGGTGAGTA
	1081	GGCTTCTCGA	GTGCCGGGAA	AAAGCTCGAG	CCTAGTTAGA	GGACTAGGAG	AGGCCGTAGC
	1141	CGTAACCTACT	CTTGGGCAAG	TAGGCGAGGC	GGTGGGTACG	CAATGGGGGC	GGCTACCTCA
	1201	GCACTAAATA	GGAGACAATT	AGACCAATTT	GAGAAAATAC	GACTTCGCCC	GAACGGAAAG
	1261	AAAAAGTACC	AAATTAAACA	TTTAATATGG	GCAGGCAAGG	AGATGGAGCG	CTTCGGCCTC
15	1321	CATGAGAGGT	TGTTGGAGAC	AGAGGAGGGG	TGTAAAAGAA	TCATAGAAGT	CCTCTACCCC
	1381	CTAGAACCAA	CAGGATCGGA	GGGCTTAAAA	AGTCTGTTCA	ATCTTGTGTG	CGTGCTATAT
	1441	TGCTTGCACA	AGGAACAGAA	AGTGAAAGAC	ACAGAGGAAG	CAGTAGCAAC	AGTAAGACAA
	1501	CACTGCCATC	TAGTGGAAAA	AGAAAAAAGT	GCAACAGAGA	CATCTAGTGG	ACAAAAAGAAA
	1561	AATGACAAGG	GAATAGCAGC	GCCACCTGGT	GGCAGTCAGA	ATTTTCCAGC	GCAACAACAA
20	1621	GGAAATGCCT	GGGTACATGT	ACCCTTGTCA	CCGCGCACCT	TAAATGCGTG	GGTAAAAGCA
	1681	GTAGAGGAGA	AAAAATTTGG	AGCAGAAATA	GTACCCATGT	TTCAAGCCCT	ATCGAATTCC
	1741	CGTTTGTGCT	AGGGTTCTTA	GGCTTCTTGG	GGGCTGCTGG	AACTGCAATG	GGAGCAGCGG
	1801	CGACAGCCCT	GACGGTCCAG	TCTCAGCATT	TGCTTGTCTG	GATACTGCAG	CAGCAGAAAG
	1861	ATCTGCTGGC	GGCTGTGGAG	GCTCAACAGC	AGATGTTGAA	GCTGACCATT	TGGGGTGTTA
25	1921	AAAACCTCAA	TGCCCCGCTC	ACAGCCCTTG	AGAAGTACCT	AGAGGATCAG	GCACGACTAA
	1981	ACTCCTGGGG	GTGCGCATGG	AAACAAGTAT	GTATACCAC	AGTGGAGTGG	CCCTGGACAA
	2041	ATCGGACTCC	GGATTGGCAA	AATATGACTT	GGTTGGAGTG	GGAAAGACAA	ATAGCTGATT
	2101	TGGAAGACAA	CATTACGAGA	CAATTAGTGA	AGGCTAGAGA	ACAAGAGGAA	AAGAATCTAG
	2161	ATGCCTATCA	GAAGTTAACT	AGTTGGTCAG	ATTCTGCTGC	TGCGTTTCAT	TTCTCAAAAT
30	2221	GGCTTAACAT	TTTAAAAATG	GGATTTTTAG	TAATAGTAGG	AATAATAGGG	TTAAGATTAC
	2281	TTTACACAGT	ATATGGATGT	ATAGTGAGGG	TTAGGCAGGG	ATATGTTTCT	CTATCTCCAC
	2341	AGATCCATAT	CCGCGGCAAT	TTTAAAAGAA	AGGGAGGAAT	AGGGGGACAG	ACTTCAGCAG
	2401	AGAGACTAAT	TAATATAATA	ACAACACAAT	TAGAAATACA	ACATTTACAA	ACCAAAATTC
	2461	AAAAAATTTT	AAATTTTAGA	GCCGCGGAGA	TCTGTTACAT	AACTTATGGT	AAATGGCCTG
35	2521	CCTGGCTGAC	TGCCCAATGA	CCCCTGCCCA	ATGATGTCAA	TAATGATGTA	TGTTCCCATG
	2581	TAATGCCAAT	AGGGACTTTC	CATTGATGTC	AATGGGTGGA	GTATTTATGG	TAAGTCCCA
	2641	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTATGCCC	CCTATTGATG	TCAATGATGG
	2701	TAAATGGCCT	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	TGGGACTTTC	CTACTTGGCA
	2761	GTACATCTAT	GTATTAGTCA	TTGCTATTAC	CATGGGAATT	CACTAGTGGG	GAAGAGCATG
40	2821	CTTGAGGGCT	GAGTGCCCTT	CAGTGGGCGT	AGAGCACATG	GCCCACAGTC	CCTGAGAAGT
	2881	TGGGGGGAGG	GGTGGGCAAT	TGAAGTGGTG	CCTAGAGAAG	GTGGGGCTTG	GGTAAACTGG
	2941	GAAAGTGATG	TGGTGTACTG	GCTCCACCTT	TTTCCCCAGG	GTGGGGGAGA	ACCATATATA
	3001	AGTGCAGTAG	TCTCTGTGAA	CATTCAAGCT	TCTGCCTTCT	CCCTCCTGTG	AGTTTGCTAG
	3061	CCACCAATGC	AGATTGAGCT	GAGCACCTGC	TTCTTCCTGT	GCCTGCTGAG	GTTCTGCTTC
45	3121	TCTGCCACCA	GGAGATACTA	CCTGGGGGCT	GTGGAGCTGA	GCTGGGACTA	CATGCAGTCT
	3181	GACCTGGGGG	AGCTGCCTGT	GGATGCCAGG	TTCCCCCCCC	GAGTGCCCAA	GAGCTTCCCC
	3241	TTCAACACCT	CTGTGGTGTA	CAAGAAGACC	CTGTTTGTGG	AGTTCACTGA	CCACCTGTTC
	3301	AACATTGCCA	AGCCCAGGCC	CCCCTGGATG	GGCCTGCTGG	GCCCCACCAT	CCAGGCTGAG
	3361	GTGTATGACA	CTGTGGTGAT	CACCTGAAG	AACATGGCCA	GCCACCCTGT	GAGCCTGCAT
50	3421	GCTGTGGGGG	TGAGCTACTG	GAAGGCCTCT	GAGGGGGCTG	AGTATGATGA	CCAGACCAGC
	3481	CAGAGGGAGA	AGGAGGATGA	CAAGGTGTTT	CCTGGGGGCA	GCCACACCTA	TGTGTGGCAG
	3541	TGCTGGAAGG	AGAATGGCCC	CATGGCCTCT	GACCCCTGTG	GCCTGACCTA	CAGTAACCTG
	3601	AGCCATGTGG	ACCTGGTGAA	GGACCTGAAC	TCTGGCCTGA	TTGGGGCCCT	GCTGGTGTGC
	3661	AGGGAGGGCA	GCCTGGCCAA	GGAGAAGACC	CAGACCCTGC	ACAAGTTTCT	CCTGCTGTTT
55	3721	GCTGTGTTTG	ATGAGGGCAA	GAGCTGGCAC	TCTGAAACCA	AGAACAGCCT	GATGCAGGAC
	3781	AGGGATGCTG	CCTCTGCCAG	GGCCTGGCCC	AAGATGCACA	CTGTGAATGG	CTATGTGAAC
	3841	AGGAGCCTGC	CTGGCCTGAT	TGGCTGCCAC	AGGAAGTCTG	TGTACTGGCA	TGTGATTGGC
	3901	ATGGGCACCA	CCCCTGAGGT	GCACAGCATC	TTCTTGAGAG	GCCACACCTT	CTGGTCAGG
	3961	AACCACAGGC	AGGCCAGCCT	GGAGATCAGC	CCCATCACCT	TCCTGACTGC	CCAGACCCTG
60	4021	CTGATGGACC	TGGGCCAGTT	CCTGCTGTTC	TGCCACATCA	GCAGCCACCA	GCATGATGGC
	4081	ATGGAGGCCT	ATGTGAAGGT	GGACAGCTGC	CCTGAGGAGC	CCCAGCTGAG	GATGAAGAAC

	4141	AATGAGGAGG	CTGAGGACTA	TGATGATGAC	CTGACTGACT	CTGAGATGGA	TGTGGTGAGG
	4201	TTTGATGATG	ACAACAGCCC	CAGCTTCATC	CAGATCAGGT	CTGTGGCCAA	GAAGCACCCC
	4261	AAGACCTGGG	TGCACTACAT	TGCTGCTGAG	GAGGAGGACT	GGGACTATGC	CCCCCTGGTG
5	4321	CTGGCCCCCTG	ATGACAGGAG	CTACAAGAGC	CAGTACCTGA	ACAATGGCCC	CCAGAGGATT
	4381	GGCAGGAAGT	ACAAGAAGGT	CAGGTTTCATG	GCCTACACTG	ATGAAACCTT	CAAGACCAGG
	4441	GAGGCCATCC	AGCATGAGTC	TGGCATCCTG	GGCCCCCTGC	TGTATGGGGA	GGTGGGGGAC
	4501	ACCCTGCTGA	TCATCTTCAA	GAACCAGGCC	AGCAGGCCCT	ACAACATCTA	CCCCCATGGC
	4561	ATCACTGATG	TGAGGCCCTT	GTACAGCAGG	AGGCTGCCCA	AGGGGGTGAA	GCACCTGAAG
10	4621	GACTTCCCCA	TCCTGCCTGG	GGAGATCTTC	AAGTACAAGT	GGACTGTGAC	TGTGGAGGAT
	4681	GGCCCCACCA	AGTCTGACCC	CAGGTGCCTG	ACCAGATACT	ACAGCAGCTT	TGTGAACATG
	4741	GAGAGGGACC	TGGCCTCTGG	CCTGATTGGC	CCCCTGCTGA	TCTGCTACAA	GGAGTCTGTG
	4801	GACCAGAGGG	GCAACCAGAT	CATGTCTGAC	AAGAGGAATG	TGATCCTGTT	CTCTGTGTTT
	4861	GATGAGAACA	GGAGCTGGTA	CCTGACTGAG	AACATCCAGA	GGTTCCTGCC	CAACCCTGCT
15	4921	GGGGTGCAGC	TGGAGGACCC	TGAGTTCAG	GCCAGCAACA	TCATGCACAG	CATCAATGGC
	4981	TATGTGTTTG	ACAGCCTGCA	GCTGTCTGTG	TGCCTGCATG	AGGTGGCCTA	CTGGTACATC
	5041	CTGAGCATTG	GGGCCCAGAC	TGACTTCCTG	TCTGTGTTCT	TCTCTGGCTA	CACCTTCAAG
	5101	CACAAGATGG	TGTATGAGGA	CACCCTGACC	CTGTTCCCCCT	TCTCTGGGGA	GACTGTGTTT
	5161	ATGAGCATGG	AGAACCCTGG	CCTGTGGATT	CTGGGCTGCC	ACAACCTCTGA	CTTCAGGAAC
20	5221	AGGGGCATGA	CTGCCCTGCT	GAAAGTCTCC	AGCTGTGACA	AGAACACTGG	GGACTACTAT
	5281	GAGGACAGCT	ATGAGGACAT	CTCTGCCTAC	CTGCTGAGCA	AGAACAATGC	CATTGAGCCC
	5341	AGGAGCTTCA	GCCAGAACAG	CAGGCACCCC	AGCACCAGGC	AGAAGCAGTT	CAATGCCACC
	5401	ACCATCCCCTG	AGAATGACAT	AGAGAAGACA	GACCCATGGT	TTGCCACCCG	GACCCCCATG
	5461	CCCAAGATCC	AGAATGTGAG	CAGCTCTGAC	CTGCTGATGC	TGCTGAGGCA	GAGCCCCACC
25	5521	CCCCATGGCC	TGAGCCTGTC	TGACCTGCAG	GAGGCCAAGT	ATGAAACCTT	CTCTGATGAC
	5581	CCCAGCCCCTG	GGGCCATTGA	CAGCAACAAC	AGCCTGTCTG	AGATGACCCA	CTTCAGGCC
	5641	CAGCTGCACC	ACTCTGGGGA	CATGGTGTTC	ACCCCTGAGT	CTGGCCTGCA	GCTGAGGCTG
	5701	AATGAGAAGC	TGGGCACCAC	TGCTGCCACT	GAGCTGAAGA	AGCTGGACTT	CAAAGTCTCC
	5761	AGCACCAGCA	ACAACCTGAT	CAGCACCATC	CCCTCTGACA	ACCTGGCTGC	TGGCACTGAC
30	5821	AACACCAGCA	GCCTGGGCCC	CCCCAGCATG	CCTGTGCACT	ATGACAGCCA	CTTGGACACC
	5881	ACCCTGTTTTG	GCAAGAAGAG	CAGCCCCCTG	ACTGAGTCTG	GGGGCCCCCT	GAGCCTGTCT
	5941	GAGGAGAACA	ATGACAGCAA	GCTGCTGGAG	TCTGGCCTGA	TGAACAGCCA	GGAGAGCAGC
	6001	TGGGGCAAGA	ATGTGAGCAG	CAGGGAGATC	ACCAGGACCA	CCCTGCAGTC	TGACCAGGAG
	6061	GAGATTGACT	ATGATGACAC	CATCTCTGTG	GAGATGAAGA	AGGAGGACTT	TGACATCTAC
35	6121	GACGAGGACG	AGAACCAGAG	CCCCAGGAGC	TTCCAGAAGA	AGACCAGGCA	CTACTTCATT
	6181	CTGTCTGTGG	AGAGGCTGTG	GGACTATGGC	ATGAGCAGCA	GCCCCCATGT	GCCCCGAAAC
	6241	AGGGCCCAGT	CTGGCTCTGT	GCCCCAGTTC	AAGAAGGTGG	TGTTCCAGGA	GTTCACTGAT
	6301	GGCAGCTTCA	CCCAGCCCCT	GTACAGAGGG	GAGCTGAATG	AGCACCTGGG	CCTGTCTGGC
	6361	CCCTACATCA	GGGCTGAGGT	GGAGACAAC	ATCATGGTGA	CCTTCAGGAA	CCAGGCCAGC
	6421	AGGCCCTACA	GCTTCTACAG	CAGCCTGATC	AGCTATGAGG	AGGACCAGAG	GCAGGGGGCT
40	6481	GAGCCCAGGA	AGAACTTTGT	GAAGCCCAAT	GAAACCAAGA	CCTACTTCTG	GAAGGTGCAG
	6541	CACCACATGG	CCCCACCAA	GGATGAGTTT	GACTGCAAGG	CCTGGGCCCTA	CTTCTCTGAT
	6601	GTGGACCTGG	AGAAGGATGT	GCACCTGCGC	CTGATTGGCC	CCCTGCTGGT	GTGCCACACC
	6661	AACACCCTGA	ACCCTGCCCA	TGGCAGGCAG	GTGACTGTGC	AGGAGTTTGC	CCTGTTCTTC
	6721	ACCATCTTTG	ATGAAACCAA	GAGCTGGTAC	TTCACTGAGA	ACATGGAGAG	GAAGTGCAGG
45	6781	GGCCCCCTGCA	ACATCCAGAT	GGAGACCCCC	ACCTTCAAGG	AGAACTACAG	GTTCCATGCC
	6841	ATCAATGGCT	ACATCATGGA	CACCCTGCCT	GGCCTGGTGA	TGGCCCAGGA	CCAGAGGATC
	6901	AGGTGGTACC	TGCTGAGCAT	GGGCAGCAAT	GAGAACATCC	ACAGCATCCA	CTTCTCTGGC
	6961	CATGTGTTCA	CTGTGAGGAA	GAAGGAGGAG	TACAAGATGG	CCCTGTACAA	CCTGTACCCCT
50	7021	GGGGTGTGTTG	AGACTGTGGA	GATGCTGCCC	AGCAAGGCTG	GCATCTGGAG	GGTGGAGTGC
	7081	CTGATTGGGG	AGCACCTGCA	TGCTGGCATG	AGCACCCCTGT	TCCTGGTGTA	CAGCAACAAG
	7141	TGCCAGACCC	CCCTGGGCAT	GGCCTCTGGC	CACATCAGGG	ACTTCCAGAT	CACTGCCTCT
	7201	GGCCAGTATG	GCCAGTGGGC	CCCCAAGCTT	GCCAGGCTGC	ACTACTCTGG	CAGCATCAAT
	7261	GCCTGGAGCA	CCAAGGAGCC	CTTCAGCTGG	ATCAAGGTGG	ACCTGCTGGC	CCCCATGATC
55	7321	ATCCATGGCA	TCAAGACCCA	GGGGGCCAGG	CAGAAGTTCA	GCAGCCTGTA	CATCAGCCAG
	7381	TTCATCATCA	TGTACAGCCT	GGATGGCAAG	AAGTGGCAGA	CCTACAGGGG	CAACAGCACT
	7441	GGCACCTTGA	TGGTGTCTT	TGGCAATGTG	GACAGCTCTG	GCATCAAGCA	CAACATCTTC
	7501	AACCCCCCA	TCATTGCCAG	ATACATCAGG	CTGCACCCCA	CCCACTACAG	CATCAGGAGC
	7561	ACCCTGAGGA	TGGAGCTGAT	GGGCTGTGAC	CTGAACAGCT	GCAGCATGCC	CTTGGGCATG
60	7621	GAGAGCAAGG	CCATCTCTGA	TGCCCAGATC	ACTGCCAGCA	GCTACTTCAC	CAACATGTTT
	7681	GCCACCTGGA	GGCCAGGCTG	CACCTGCAGG	GCAGGAGCAA	TGCCTGGAGG	
	7741	CCCCAGGTCA	ACAACCCCAA	GGAGTGGCTG	CAGGTGGACT	TCCAGAAGAC	CATGAAGGTG

	7801	ACTGGGGTGA	CCACCCAGGG	GGTGAAGAGC	CTGCTGACCA	GCATGTATGT	GAAGGAGTTC
	7861	CTGATCAGCA	GCAGCCAGGA	TGGCCACCAG	TGGACCCTGT	TCTTCCAGAA	TGGCAAGGTG
	7921	AAGGTGTTCC	AGGGCAACCA	GGACAGCTTC	ACCCCTGTGG	TGAACAGCCT	GGACCCCCCC
	7981	CTGCTGACCA	GATACCTGAG	GATTCACCCC	CAGAGCTGGG	TGCACCAGAT	TGCCCTGAGG
5	8041	ATGGAGGTGC	TGGGCTGTGA	GGCCAGGAC	CTGTACTGAG	CGGCCGCGGG	CCCAATCAAC
	8101	CTCTGGATTA	CAAAATTGT	AAAAGATTGA	CTGGTATTCT	TAACATATGT	GCTCCTTTTA
	8161	CGCTATGTGG	ATACGCTGCT	TTAATGCCTT	TGTATCATGC	TATTGCTTCC	CGTATGGCTT
	8221	TCATTTTCTC	CTCCTTGTAT	AAATCCTGGT	TGCTGTCTCT	TTATGAGGAG	TTGTGGCCCC
	8281	TTGTCAAGCA	ACGTGGCGTG	GTGTGCACTG	TGTTTGCTGA	CGCAACCCCC	ACTGGTTGGG
10	8341	GCATTGCCAC	CACCTGTGAG	CTCCTTTCCG	GGACTTTCCG	TTTCCCCCTC	CCTATTGCCA
	8401	CGGCGGAAC	CATCGCCGCC	TGCCCTTGCC	GCTGCTGGAC	AGGGGCTCGG	CTGTTGGGCA
	8461	CTGACAATT	CGTGGTGTG	TGGGGGAAAT	CATCGTCCTT	TCCTTGGCTG	CTCGCCTGTG
	8521	TTGCCACCTG	GATTCTGCGC	GGGACGTCTT	TCGTCTACGT	CCCTTCGGCC	CTCAATCCAG
	8581	CGGACCTTCC	TTCCCGCGGC	CTGTGCGCGG	CTCTGCGGCC	TCTTCCGCGT	CTTCGCCTTC
15	8641	GGCCTCAGAC	GAGTCGGATC	TCCCTTTGGG	CCGCCTCCCC	GCAAGCTTCG	CACTTTTTAA
	8701	AAGAAAAGGG	AGGACTGGAT	GGGATTTATT	ACTCCGATAG	GACGCTGGCT	TGTAACTCAG
	8761	TCTCTTACTA	GGAGACCAGC	TTGAGCCTGG	GTGTTTCGCTG	GTTAGCCTAA	CTGGTTGGC
	8821	CACCAGGGGT	AAGGACTCCT	TGGCTTAGAA	AGCTAATAAA	CTTGCCTGCA	TTAGAGCTCT
	8881	TACGCGTCCC	GGGCTCGAGA	TCCGCATCTC	AATTAGTCAG	CAACCATAGT	CCCGCCCCCTA
20	8941	ACTCCGCCCA	TCCCGCCCTT	AACTCCGCC	AGTCCGCC	ATTCTCCGCC	CCATGGCTGA
	9001	CTAATTTTTT	TTATTTATGC	AGAGCCGAG	GCCGCTCGG	CCTCTGAGCT	ATTCCAGAAG
	9061	TAGTGAGGAG	GCTTTTTTGG	AGGCCTAGGC	TTTTGCAAAA	AGCTAAGTTC	TTTATTGCAG
	9121	CTTATAATGG	TTACAAATAA	AGCAATAGCA	TCACAAATTT	CACAAATAAA	GCATTTTTTT
	9181	CACATGCATC	TAGTTGTGGT	TTGTCCAAAC	TCATCAATGT	ATCTTATCAT	GTCTGTCCGC
25	9241	TTCTTCGCTC	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA
	9301	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG
	9361	AGCAAAAGGC	CAGCAAAAGG	CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA
	9421	TAGGCTCCGC	CCCCCTGACG	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA
	9481	CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC
30	9541	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAGCGTGGC
	9601	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTT	GCTCCAAGCT
	9661	GGGCTGTGTG	CACGAACCCC	CCGTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACATATC
	9721	TCTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG
	9781	GATTAGCAGA	GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA
35	9841	CGGCTACACT	AGAAGAACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG
	9901	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT
	9961	TGTTTGCAAG	CAGCAGATTA	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT
	10021	TTCTACGGGG	TCTGACGCTC	AGTGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG
40	10081	ATTATCAAAA	AGGATCTTCA	CCTAGATCCT	TTTAAATTA	AAATGAAGTT	TTAAATCAAT
	10141	CTAAAGTATA	TATGAGTAAA	CTTGGTCTGA	CAGTTAGAAA	AACTCATCGA	GCATCAAAATG
	10201	AAATGCAAT	TTATTCATAT	CAGGATTATC	AATACCATAT	TTTTGAAAAA	GCCGTTTCTG
	10261	TAATGAAGGA	GAAAACTCAC	CGAGGCAGTT	CCATAGGATG	GCAAGATCCT	GGTATCGGTC
	10321	TGCGATTCCG	ACTCGTCCAA	CATCAATACA	ACCTATTAAT	TTCCCTCTCG	CAAAAAATAAG
	10381	GTTATCAAGT	GAGAAATCAC	CATGAGTGAC	GACTGAATCC	GGTGAGAATG	GCAACAGCTT
45	10441	ATGCATTTCT	TTCCAGACTT	GTTCAACAGG	CCAGCCATTA	CGCTCGTCAT	CAAAATCACT
	10501	CGCATCAACC	AAACCGTTAT	TCATTCGTGA	TTGCGCCTGA	GCGAGACGAA	ATACGCGATC
	10561	GCTGTTAAAA	GGACAATTAC	AAACAGGAAT	CGAATGCAAC	CGGCGCAGGA	ACACTGCCAG
	10621	CGCATCAACA	ATATTTTCAC	CTGAATCAGG	ATATTCTTCT	AATACCTGGA	ATGCTGTTTT
	10681	TCCGGGGATC	GCAGTGGTGA	GTAACCATGC	ATCATCAGGA	GTACGGATAA	AATGCTTGAT
50	10741	GGTCGGAAGA	GGCATAAATT	CCGTCAGCCA	GTTTAGTCTG	ACCATCTCAT	CTGTAACATC
	10801	ATTGGCAACG	CTACCTTTGC	CATGTTTCAG	AAACAACCTCT	GGCGCATCGG	GCTTCCCATC
	10861	CAATCGATAG	ATTGTCGCAC	CTGATTGCC	GACATTATCG	CGAGCCCAT	TATACCCATA
	10921	TAAATCAGCA	TCCATGTTGG	AATTTAATCG	CGGCCTAGAG	CAAGACGTTT	CCCGTTGAAT
	10981	ATGGCTCATA	ACACCCCTTG	TATTACTGTT	TATGTAAGCA	GACAGTTTTA	TTGTTTCATGA
55	11041	TGATATATTT	TTATCTTGTG	CAATGTAACA	TCAGAGATTT	TGAGACACAA	CAATTGGTCC
	11101	ACGGATCC					

SEQ ID NO: 29 Exemplary CAG promoter

Length: 1738; Molecule Type: DNA; Features Location/Qualifiers: source, 1..1738; mol_type, other DNA; note, CAG promoter; organism, synthetic construct

```
ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCG
5 TTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATGACGT
ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT
TGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGC
ATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
10 TGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACCCCAATTTTGTATTAT
TTATTTTTTAAATTATTTTGTGCAGCGATGGGGCGGGGGGGGGGGGGGGCGCGCCAGGCGGGGCGGGGCGGG
GCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTT
TTATGGCGAGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGCGCGCTGCC
TTCGCCCCGTGCCCCGCTCCGCGCCCGCTCGCGCCCGCCGCCCCGGCTCTGACTGACCGCTTACTCCACAGG
15 TGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAAATGACGGCTTGTTCCTTTCTGT
GGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGAGCGGCTCGGGGGGTGCGTGCGTGT
GTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCGCGGCTGTGAGCGCTGCGGGCGCGGCGGGGC
TTTGTGCGCTCCGCAGTGTGCGGAGGGGAGCGCGCCGGGGCGGTGCCCGCGGTGCGGGGGGGCTGCCAGG
GGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGCGGGGGTGAGCAGGGGTGTGGGCGCGTCGGTCGGGCTGCAAC
20 CCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCG
CGGGGCTCGCCGTGCCGGGCGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCCTCGGGCCGGG
AGGGCTCGGGGAGGGGCGGGCGGCCCGGAGCGCGGCGGCTGTGAGGCGCGGCGAGCCGAGCCATTGCC
TTTTATGGTAATCGTGCGAGAGGGCGCAGGACTTCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGC
GCCGCCGCACCCCTCTAGCGGGCGGGGCGAAGCGGTGCGGCGCGGCGAGGAAGGAAATGGGCGGGGAGGGCC
25 TTCGTGCGTCGCCGCGCCGCTCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGACGGCTGCCTTC
GGGGGGGACGGGGCAGGGCGGGGTTTCGGCTTCTGGCGTGTGACCGCGGGCTCTAGAGCCTCTGCTAACCATGTT
ATGCCTTCTTCTTTTCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAAT
TGCTCGAGCCACC
```

CLAIMS

1. A method of producing a retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, and which comprises a promoter and a transgene, wherein said method comprises the use of codon-optimised gag-pol genes.
5
2. The method of claim 1, wherein the retroviral vector is a lentiviral vector.
- 10 3. The method of claim 2, wherein the lentiviral vector is selected from the group consisting of a Simian immunodeficiency virus (SIV) vector, a Human immunodeficiency virus (HIV) vector, a Feline immunodeficiency virus (FIV) vector, an Equine infectious anaemia virus (EIAV) vector, and a Visna/maedi virus vector.
- 15 4. The method of claim 2 or 3, wherein the lentiviral vector is an SIV vector.
5. The method of any one of the preceding claims, wherein the codon-optimised gag-pol genes are SIV gag-pol genes.
- 20 6. The method of any one of the preceding claims, wherein the codon-optimised gag-pol genes comprise or consist of a nucleic acid sequence having at least 80% sequence identity to SEQ ID NO: 1.
7. The method of claim 6, wherein the codon-optimised gag-pol genes comprise or consist of the nucleic acid sequence of SEQ ID NO: 1.
25
8. The method of any one of the preceding claims, wherein the codon-optimised gag-pol genes are comprised in a plasmid that comprises or consists of a nucleic acid sequence having at least 80% sequence identity to SEQ ID NO: 5.
30
9. The method of claim 8, wherein the codon-optimised gag-pol genes are comprised in a plasmid that comprises or consists of the nucleic acid sequence of SEQ ID NO: 5.

10. The method of any one of the preceding claims, wherein the respiratory paramyxovirus is a Sendai virus.
11. The method of any one of the preceding claims, wherein the titre of retroviral vector produced is:
- a) equivalent to the titre of retroviral vector produced by a corresponding method which does not use codon-optimised gal-pol genes; or
 - b) increased compared with the titre of retroviral vector produced by a corresponding method which does not use codon-optimised gal-pol genes.
12. The method of claim 11, wherein the titre of retroviral vector is at least 2-fold, or at least 2.5-fold greater than the titre of retroviral vector produced by a corresponding method which does not use codon-optimised gal-pol genes.
13. The method of any one of the preceding claims, wherein the promoter is selected the group consisting of a cytomegalovirus (CMV) promoter, elongation factor 1a (EF1a) promoter, and a hybrid human CMV enhancer/EF1a (hCEF) promoter.
14. The method of any one of the preceding claims, wherein the vector comprises a hybrid human CMV enhancer/EF1a (hCEF) promoter.
15. The method of any one of the preceding claims, wherein the transgene is selected from:
- a) a secreted therapeutic protein, optionally Alpha-1 Antitrypsin (A1AT), Factor VIII, Surfactant Protein B (SFTPB), Factor VII, Factor IX, Factor X, Factor XI, von Willebrand Factor, Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) and a monoclonal antibody against an infectious agent; or
 - b) *CFTR*, *ABCA3*, *DNAH5*, *DNAH11*, *DNAI1*, and *DNAI2*.
16. The method of any one of the preceding claims, wherein the transgene encodes:
- a) *CFTR*;
 - b) A1AT; or
 - c) FVIII.
17. The method of any one of the preceding claims, wherein:
- a) the promoter is a hCEF promoter and the transgene encodes *CFTR*;

- b) the promoter is a hCEF promoter and the transgene encodes A1AT; or
- c) the promoter is a hCEF or CMV promoter and the transgene encodes FVIII.

18. The method of any one of the preceding claims, said method comprising the following steps:
- a) growing cells in suspension;
 - b) transfecting the cells with one or more plasmids;
 - c) adding a nuclease;
 - d) harvesting the lentivirus;
 - e) adding trypsin; and
 - f) purification.
19. The method according to claim 18, wherein the one or more plasmids comprise or consist of:
- a) a vector genome plasmid, preferably selected from selected from pGM830 and pGM326;
 - b) a co-galpol plasmid, preferably pGM691;
 - c) a Rev plasmid, preferably pGM299;
 - d) a fusion (F) protein plasmid, preferably pGM301; and
 - e) a hemagglutinin-neuraminidase (HN) plasmid, preferably pGM303.
20. The method according to claim 19, wherein the ratio of vector genome plasmid: co-gagpol plasmid: Rev plasmid: F plasmid: HN plasmid is 20:9:6:6:6.
21. The method according to any one of claims 18 to 20, wherein steps (a)-(f) are carried out sequentially.
22. The method according to any one of claims 18 to 21, wherein the cells are HEK293T or 293T/17 cells.
23. The method according to any one of claims 18 to 22, wherein the addition of the nuclease is at the pre-harvest stage.
24. The method according to any one of claims 18 to 23, wherein the addition of trypsin is at the post-harvest stage.

25. The method according to any one of claims 18 to 24, wherein the purification step comprises a chromatography step.
26. The method according to any one of the claims 19 to 24, wherein the vector genome plasmid is modified to reduce the number of retroviral ORFs.
27. A nucleic acid comprising codon-optimised gag-pol genes, said nucleic acid having at least 80% sequence identity to SEQ ID NO: 1.
28. The nucleic acid of claim 27 which comprises or consists of the nucleic acid sequence of SEQ ID NO: 1.
29. A plasmid comprising a nucleic acid as defined in claim 27 or 28, wherein optionally:
- a) the plasmid comprises or consists of a nucleic acid sequence having at least 80% sequence identity to SEQ ID NO: 5; or
 - b) the plasmid comprises or consists of the nucleic acid sequence of SEQ ID NO: 5.
30. A host cell comprising a nucleic acid as defined in claim 27 or 28, and/or a plasmid as defined in claim 29.
31. A retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus which is obtainable by a method as defined in any one of claims 1 to 26.
32. A method of treating a disease comprising administering a retroviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus which is obtainable by a method as defined in any one of claims 1 to 26, to a subject in need thereof.
33. The method of treatment according to claim 32, wherein the disease to be treated is a lung disease, preferably cystic fibrosis.