

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(12)

Oversættelse af europæisk patentskrift

(10) **DK/EP 2194137 T3**

-
- (51) Int.Cl.: *C 12 N 7/01 (2006.01)* *A 61 K 48/00 (2006.01)* *C 07 K 14/155 (2006.01)*
C 07 K 14/16 (2006.01) *C 12 N 5/10 (2006.01)* *C 12 N 7/02 (2006.01)*
C 12 N 7/04 (2006.01) *C 12 N 15/49 (2006.01)* *C 12 N 15/867 (2006.01)*
- (45) Oversættelsen bekendtgjort den: **2017-03-20**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2016-12-21**
- (86) Europæisk ansøgning nr.: **10158183.3**
- (86) Europæisk indleveringsdag: **2001-04-18**
- (87) Den europæiske ansøgnings publiceringsdag: **2010-06-09**
- (30) Prioritet: **2000-04-19 GB 0009760**
- (62) Stamansøgningsnr: **01921651.4**
- (84) Designerede stater: **AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR**
- (73) Patenthaver: **Oxford BioMedica (UK) Limited, Windrush Court , Transport Way, Oxford, OX4 6LT, Storbritannien**
- (72) Opfinder: **Kingsman, Alan John, Oxford BioMedica (UK) Limited, Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, , Oxford OX4 4GA, Storbritannien**
Kim, Narry, Oxford BioMedica (UK) Limited, Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA, Storbritannien
Kotsopoulou, Ekaterini, Oxford BioMedica (UK) Limited, Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA, Storbritannien
Rohli, Jonathan, Oxford BioMedica (UK) Limited, Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA, Storbritannien
Mitrophanous, Kyriacos, Oxford BioMedica (UK) Limited, Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA, Storbritannien
- (74) Fuldmægtig i Danmark: **RWS Group, Europa House, Chiltern Park, Chiltern Hill, Chalfont St Peter, Bucks SL9 9FG, Storbritannien**
- (54) Benævnelse: **Celler, der omfatter kodon-optimerede lentiviruspartikler**
- (56) Fremdragne publikationer:
WO-A-00/15819
WO-A-00/39302
WO-A-00/65076
WO-A-99/32646
WO-A-99/41397
WO-A1-00/55341
J. ZUR MEGEDE ET AL.: "Increased expression and immunogenicity of sequence-modified HIV-1 gag gene" J. VIROL., vol. 74, no. 6, March 2000 (2000-03), pages 2628-2635, XP000980650
E. KOTSOPOLLOU ET AL.: "A rev-independent HIV 1-based vector that exploits a codon optimised HIV-1 gag-pol gene" J. VIROL., vol. 74, no. 10, 2000, pages 4839-4852, XP002140792

Fortsættes ...

R. WAGNER ET AL.: "Rev-independent expression of synthetic gag-pol genes of HIV-1 and SIV: implications for the safety of lentiviral vectors" HUMAN GENE THERAPY, vol. 11, 2000, pages 2403-2413, XP001064159

G. BUCHSCHACHER ET AL.: "Development of lentiviral vectors for gene therapy of human diseases" BLOOD, vol. 95, no. 6, 15 April 2000 (2000-04-15), pages 2499-2504, XP000916373

CHUNG ET AL.: "Development of new safer retrovirus packaging sequences and cell lines" J. CELL. BIOL. SUPPLEMENT, vol. 0, no. 21A, 1995, page 404, XP001064207

ZUFFEREY R ET AL: "SELF-INACTIVATING LENTIVIRUS VECTOR FOR SAFE AND EFFICIENT IN VIVO GENE DELIVERY", JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 72, no. 12, 1 December 1998 (1998-12-01), pages 9873-9880, XP002929451, ISSN: 0022-538X

DESCRIPTION

Field of the Invention

[0001] The present invention relates to methods of improving the safety of retroviral vectors capable of delivering therapeutic genes for use in gene therapy, and to novel nucleotide sequences for use in such methods.

Background to the Invention

[0002] Retroviral vectors are now widely used as vehicles to deliver genes into cells. Their popularity stems from the fact that they are easy to produce and mediate stable integration of the gene that they carry into the genome of the target cell. This enables long-term expression of the delivered gene (1).

[0003] There has been considerable interest, for some time, in the development of retroviral vector systems based on lentiviruses. Lentiviruses are a small subgroup of complex retroviruses. They contain, in addition to the common retroviral genes (*gag*, *pol* and *env*), genes which enable them to regulate their life cycle and to infect non-dividing cells (2). Vector systems based thereon are therefore of interest because of their potential use in the transfer of a gene of interest to non-dividing cells such as neurones. In addition, lentiviral vectors enable very stable long-term expression of the gene of interest. This has been shown to be at least three months for transduced rat neuronal cells while MLV based vectors were only able to express the gene of interest for six weeks.

[0004] The most commonly used lentivirus is the Human Immunodeficiency Virus (HIV), the etiologic agent of AIDS (acquired immune deficiency syndrome). HIV-based vectors have been shown to efficiently transduce non-dividing cells (3) and can be used, for example, to target anti-HIV therapeutic genes to HIV susceptible cells. However, HIV vectors have a number of significant disadvantages that may limit their therapeutic application to certain diseases. In particular, HIV-1 is a human pathogen carrying potentially oncogenic proteins and sequences. There is the risk that introduction of vector particles produced in packaging cells which express HIV *gag-pol* will introduce these proteins into the patient leading to seroconversion.

[0005] Emphasis has therefore been placed on the safety of these vectors. One strategy looks at the design of production systems for retroviral vectors. A retrovirus vector system basically consists of two elements, a packaging cell line and a vector genome. The simplest packaging line consists of a provirus in which the ψ sequence (a determinant of RNA packaging reporting in HIV as lying between U5 and *gag*) has been deleted. When stably transfected into a cell, virus particles containing reverse transcriptase will be produced but virion RNA will not become packaged within these particles. The complementing component in a retrovirus vector system is the genome vector itself. The genome vector needs to contain a packaging sequence but much of the structural coding regions can be deleted. Often a selectable marker gene, or other nucleotide sequence of interest, is incorporated into the vector. Vector stocks of the packaging line can then be used to infect target cells. Provided the cell is successfully infected by the viral particle, the genome vector sequence will be reverse transcribed and integrated by the retroviral machinery. However, infection is an end process so no further replication or spread of the vector should occur.

[0006] As indicated above, however, problems are encountered in the design of safe and effective retroviral vectors. These include the possibility that recombination between the packaging vector and the packaging sequence can lead to the generation of wild type replication competent virus. Consequently efforts have been directed at improving the safety of packaging cell constructs.

[0007] In second generation packaging cell lines, in addition to deletion of the packaging sequence, the 3' LTR was also deleted so that two recombinations are necessary to generate a wild type virus.

[0008] In third generation packaging lines the *gag-pol* genes and *env* gene are placed on separate constructs that are sequentially introduced into the packaging cells to prevent recombination during transfection.

[0009] With regard to the packaging signal, EP 0 368 882A (Sodroski) discloses that in HIV it corresponds to the region between the 5' major splice donor and the *gag* initiation codon, and particularly corresponds to a segment just downstream of the 5' major splice donor, and about 14 bases upstream of the *gag* initiation codon. It is this region which Sodroski teaches should be deleted from the *gag-pol* cassette. WO97/12622 (Verma) describes that in HIV-1 a 39 bp internal deletion in the ψ sequence can be made between the 5' splice donor site and the starting codon of the *gag* gene.

[0010] Codon wobbling can be used to reduce recombination frequency while maintaining the primary protein sequence of the constructs, c.f. (4) in which the region of overlap between the *gag-pol* and *env* expression constructs was reduced to 61 bp extending over the common region between *pol* and *env* which are in different reading frames. Transversion mutations were introduced into the final 20 codons of *pol*, retaining the integrity of the coding region while reducing the homology with *env* to 55% in the overlap region. Similarly wobble mutations were introduced into the 3' of *env* and all sequences downstream of the *env* stop codon were deleted.

[0011] Efficient vectors usually contain part of *gag* on the genome vector to increase virion titre. Unlike the packaging sequence which can be in any position within a sequence to effect packaging, the *gag* sequence must be in its native position adjacent to ψ to have any effect.

[0012] It will be appreciated that whilst significant improvements in packaging cell and vector design have been made there is still scope for further refinement of current packaging lines.

[0013] WO 00/55341 and WO 99/41397 relate to viral vectors capable of delivering anti-viral inhibitory RNA molecules to target cells, particular for the treatment of AIDS and HIV infection.

[0014] WO 00/15819 relates to packaging cell lines for producing a viral accessory protein independent lentivirus-derived retroviral particle which comprises a coding sequence for lentivirus *gagpol* which has been codon optimised. Zufferey R et al., Journal of Virology, Dec 1998, p. 9873-9880 which describes self-inactivating lentivirus vectors for safe and efficient *in vivo* gene delivery.

Summary of the Invention

[0015] It is therefore an aim of the invention to provide retroviral particles, in particular lentiviral particles, and particularly those which carry nucleotide constructs encoding therapeutic proteins, that have improved safety over the corresponding wild type viral particle. In our WO99/41397 we describe codon optimisation of the *gag-pol* genes as a means of overcoming the Rev/RRE requirement for export and to enhance RNA stability. We have now found however that the codon optimised *gag-pol* sequence overcomes potential recombination problems with vector genomes which carry part of a *gag* sequence with the aim of increasing titre. This strategy also avoids the need to use *gag* regions from different viruses in the packaging and vector genome constructs.

[0016] Another significant advantage provided by the invention is that the codon optimisation disrupts RNA secondary structures, such as the packaging signal, thus rendering the *gag-pol* mRNA non-packagable. Thus, the present invention allows retroviral sequence upstream of the *gag* initiation codon to be retained, in contrast to Sodroski and Verma, without significantly compromising safety.

Statements of the Invention

[0017] According to one aspect of the present invention there is provided a method of producing a replication defective HIV vector comprising transfecting a producer cell with the following:

1. i) an HIV genome comprising a nucleotide of interest (NOI), a packaging signal, and RRE;
2. ii) a nucleotide sequence coding for HIV gag and pol proteins; and
3. iii) a nucleotide sequence coding for an env protein; characterised in that the nucleotide sequence coding for the HIV gag and pol proteins is codon optimised for expression in the producer cell with the exception of the sequence encompassing the frameshift site and wherein (ii) does not comprise RRE; wherein the producer cell expresses *rev* and wherein the HIV vector is a self-inactivating vector.

[0018] According to another aspect of the present invention there is provided an HIV vector particle producer cell comprising:

1. i) an HIV genome comprising at least one nucleotide sequence of interest, packaging signal, and RRE; and
2. ii) a nucleotide sequence coding for HIV gag and pol proteins; and
3. iii) a nucleotide sequence coding for an env protein; characterised in that the nucleotide sequence coding for the HIV gag

and pol proteins is codon optimised for expression in the producer cell with the exception of the sequence encompassing the frameshift site and wherein (ii) does not comprise RRE; wherein the producer cell expresses *rev* and wherein the HIV vector is a self-inactivating vector.

[0019] We also provide novel codon optimised sequences as shown in SEQ ID NOS: 15 which may be used in the present invention. However, it will be appreciated that any convenient codon optimised *gag-pol* sequence may be employed in the invention.

[0020] The present invention also provides a pharmaceutical composition comprising a producer cell according to the present invention, together with a pharmaceutically acceptable diluent or carrier.

[0021] By "reducing" we mean that the chance of an event occurring is reduced compared to a comparable population having the wild-type *gag-pol* sequence. Within a population the chance of an event occurring may be prevented for an individual retrovirus vector or particle.

Detailed Description of the Invention

[0022] Various preferred features and embodiments of the present invention will now be described by way of non-limiting example.

[0023] The present invention employs the concept of codon optimisation.

[0024] Codon optimisation has previously been described in our WO99/41397 as a means of overcoming the Rev/RRE requirement for export and to enhance RNA stability. The alterations to the coding sequences for the viral components improve the sequences for codon usage in the mammalian cells or other cells which are to act as the producer cells for retroviral vector particle production. This improvement in codon usage is referred to as "codon optimisation". Many viruses, including HIV and other lentiviruses, use a large number of rare codons and by changing these to correspond to commonly used mammalian codons, increased expression of the packaging components in mammalian producer cells can be achieved. Codon usage tables are known in the art for mammalian cells, as well as for a variety of other organisms.

[0025] By virtue of alterations in their sequences, the nucleotide sequences encoding the packaging components of the viral particles required for assembly of viral particles in the producer cells/packaging cells have RNA instability sequences (INS) eliminated from them. At the same time, the amino acid coding sequence for the packaging components is retained so that the viral components encoded by the sequences remain the same, or at least sufficiently similar that the function of the packaging components is not compromised.

[0026] The term "viral polypeptide required for the assembly of viral particles" means a polypeptide normally encoded by the viral genome to be packaged into viral particles, in the absence of which the viral genome cannot be packaged. For example, in the context of retroviruses such polypeptides would include *gag-pol* and *env*. The term "packaging component" is also included within this definition.

[0027] As discussed in our WO99/32646, the sequence requirements for packaging HIV vector genomes are complex. The HIV-1 packaging signal encompasses the splice donor site and contains a portion of the 5'-untranslated region of the *gag* gene, which has a putative secondary structure containing 4 short stem-loops. However, additional sequences elsewhere in the genome are also known to be important for efficient encapsidation of HIV. For example, the first 350 bps of the *gag* protein coding sequence may contribute to efficient packaging. Thus, for construction of HIV-1 vectors capable of expressing heterologous genes, a packaging signal extending to 350 bps of the *gag* protein-coding region has been used on the vector genome. We have now found that codon optimisation of the *gag* coding region on the packaging vector, at least in the region into which the packaging signal extends, also has the effect of disrupting packaging of the vector genome. Thus codon optimisation is a novel method of obtaining a replication defective viral particle.

[0028] Also as disclosed in WO99/32646, the structure of the packaging signal in equine lentiviruses is different from that of HIV. Instead of a short sequence of 4 stem loops together with a packaging signal extending to 350 bps of the *gag* protein-coding region, we have found that in equine lentiviruses the packaging signal may not extend as far into the *gag* protein-coding region as may have been thought.

[0029] In one aspect only codons relating to the packaging signal are codon optimised. Thus, in one aspect, codon optimisation extends to at least the first 350 bps of the *gag* protein coding region. In equine lentiviruses, at least, codon optimisation extends to at least nucleotide 300 of the *gag* coding region, more preferably to at least nucleotide 150 of the *gag* coding region. Although not optimal, codon optimisation could extend to, say, only the first 109 nucleotides of the *gag* coding region. It may also be possible for codon optimisation to extend to only the first codon of the *gag* coding region.

[0030] However, in the present invention the sequences are codon optimised in their entirety, with the exception of the sequence encompassing the frameshift site. The *gag-pol* gene comprises two overlapping reading frames encoding *gag* and *pol* proteins respectively. The expression of both proteins depends on a frameshift during translation. This frameshift occurs as a result of ribosome "slippage" during translation. This slippage is thought to be caused at least in part by ribosome-stalling RNA secondary structures. Such secondary structures exist downstream of the frameshift site in the *gag-pol* gene. For HIV, the region of overlap extends from nucleotide 1222 downstream of the beginning of *gag* (wherein nucleotide 1 is the A of the *gag* ATG) to the end of *gag* (nt 1503). Consequently, a 281 bp fragment spanning the frameshift site and the overlapping region of the two reading frames is preferably not codon optimised. Retaining this fragment will enable more efficient expression of the *gag-pol* proteins.

[0031] Derivations from optimal codon usage may be made, for example, in order to accommodate convenient restriction sites, and conservative amino acid changes may be introduced into the *gag-pol* proteins.

[0032] In a highly preferred embodiment, codon optimisation was based on lightly expressed mammalian genes. The third and sometimes the second and third base may be changed. An example of a codon usage table is given in Figure 3b.

[0033] Due to the degenerate nature of the Genetic Code, it will be appreciated that numerous *gag-pol* sequences can be achieved by a skilled worker. Also there are many retroviral variants described and which can be used as a starting point for generating a codon optimised *gag-pol* sequence. Lentiviral genomes can be quite variable. For example there are many quasi-species of HIV-1 which are still functional. This is also the case for EIAV. These variants may be used to enhance particular parts of the transduction process. Examples of HIV-1 variants may be found at <http://hiv-web.lanl.gov>. Details of EIAV clones may be found at the NCBI database: <http://www.ncbi.nlm.nih.gov>.

[0034] As codon optimisation may result in disruption of RNA secondary structures such as the packaging signal, it will be appreciated that any endogenous packaging signal upstream of the *gag* initiation codon could be retained without compromising safety.

[0035] An additional advantage of codon optimising packaging components is that this can increase gene expression. In particular, it can render *gag-pol* expression Rev independent. In order to enable the use of anti-*rev* or RRE factors in the retroviral vector, however, it would be necessary to render the viral vector generation system totally Rev/RRE independent (5). Thus, the genome also needs to be modified. This is achieved by optimising vector genome components. Advantageously, these modifications also lead to the production of a safer system absent of all accessory proteins both in the producer and in the transduced cell, and are described below.

[0036] As described above, the packaging components for a retroviral vector include expression products of *gag*, *pol* and *env* genes. In addition, efficient packaging depends on a short sequence of 4 stem loops followed by a partial sequence from *gag* and *env* (the "packaging signal"). Thus, inclusion of a deleted *gag* sequence in the retroviral vector genome (in addition to the full *gag* sequence on the packaging construct) will optimise vector titre. To date efficient packaging has been reported to require from 255 to 360 nucleotides of *gag* in vectors that still retain *env* sequences, or about 40 nucleotides of *gag* in a particular combination of splice donor mutation, *gag* and *env* deletions. We have surprisingly found that a deletion of up to 360 nucleotides in *gag* leads to an increase in vector titre. Further deletions resulted in lower titres. Additional mutations at the major splice donor site upstream of *gag* were found to disrupt packaging signal secondary structure and therefore lead to decreased vector titre. Thus, preferably, the retroviral vector genome includes a *gag* sequence from which up to 360 nucleotides have been removed.

[0037] We therefore allow the preparation of a so-called "minimal" system in which all of the accessory genes may be removed. In HIV these accessory genes, are *vpr*, *vif*, *tat*, *nef*, *vpu* and *rev*. Similarly, in other lentiviruses the analogous accessory genes normally present in the lentivirus may be removed. For the avoidance of doubt, however, we would mention that the present invention also extends to systems, particles and vectors in which one or more of these accessory genes is present and in any combination.

[0038] The term "viral vector" refers to a nucleotide construct comprising a viral genome capable of being transcribed in a host

cell, which genome comprises sufficient viral genetic information to allow packaging of the viral RNA genome, in the presence of packaging components, into a viral particle capable of infecting a target cell. Infection of the target cell includes reverse transcription and integration into the target cell genome, where appropriate for particular viruses. The viral vector in use typically carries heterologous coding sequences (nucleotides of interest or "NOIs") which are to be delivered by the vector to the target cell, for example a first nucleotide sequence encoding a ribozyme. By "replication defective" we mean that a viral vector is incapable of independent replication to produce infectious viral particles within the final target cell.

[0039] The term "viral vector system" is intended to mean a kit of parts which can be used when combined with other necessary components for viral particle production to produce viral particles in host cells. For example, an NOI may typically be present in a plasmid vector construct suitable for cloning the NOI into a viral genome vector construct. When combined in a kit with a further nucleotide sequence, which will also typically be present in a separate plasmid vector construct, the resulting combination of plasmid containing the NOI and plasmid containing the further nucleotide sequence comprises the essential elements of the invention. Such a kit may then be used by the skilled person in the production of suitable viral vector genome constructs which when transfected into a host cell together with the plasmid containing the further nucleotide sequence, and optionally nucleic acid constructs encoding other components, required for viral assembly, will lead to the production of infectious viral particles.

[0040] Alternatively, the further nucleotide sequence may be stably present within a packaging cell line that is included in the kit.

[0041] The kit may include the other components needed to produce viral particles, such as host cells and other plasmids encoding essential viral polypeptides required for viral assembly. By way of example, the kit may contain (i) a plasmid containing an NOI and (ii) a plasmid containing a further nucleotide sequence encoding a modified retroviral *gag-pol* construct which has been codon optimised for expression in a producer of choice. Optional components would then be (a) a retroviral genome construct with suitable restriction enzyme recognition sites for cloning the NOI into the viral genome, optionally with at least a partial *gag* sequence; (b) a plasmid encoding a VSV-G env protein. Alternatively, nucleotide sequence encoding viral polypeptides required for assembly of viral particles may be provided in the kit as packaging cell lines comprising the nucleotide sequences, for example a VSV-G expressing cell line.

[0042] The term "viral vector production system" refers to the viral vector system described above wherein the NOI has already been inserted into a suitable viral vector genome.

[0043] In the present invention, several terms are used interchangeably. Thus, "virion", "virus", "viral, particle", "retroviral particle", "retrovirus", and "vector particle" mean virus and virus-like particles that are capable of introducing a nucleic acid into a cell through a viral-like entry mechanism. Such vector particles can, under certain circumstances, mediate the transfer of NOIs into the cells they infect. A retrovirus is capable of reverse transcribing its genetic material into DNA and incorporating this genetic material into a target cell's DNA upon transduction. Such cells are designated herein as "target cells".

[0044] As used herein the term "target cell" simply refers to a cell which the regulated retroviral vector of the present disclosure, whether native or targeted, is capable of infecting or transducing.

[0045] A lentiviral vector particle for use according to the invention will be capable of transducing cells which are slowly-dividing, and which non-lentiviruses such as MLV would not be able to efficiently transduce. Slowly-dividing cells divide once in about every three to four days including certain tumour cells. Although tumours contain rapidly dividing cells, some tumour cells especially those in the centre of the tumour, divide infrequently.

[0046] Alternatively the target cell may be a growth-arrested cell capable of undergoing cell division such as a cell in a central portion of a tumour mass or a stem cell such as a haematopoietic stem cell or a CD34-positive cell.

[0047] As a further alternative, the target cell may be a precursor of a differentiated cell such as a monocyte precursor, a CD33-positive cell, or a myeloid precursor.

[0048] As a further alternative, the target cell may be a differentiated cell such as a neuron, astrocyte, glial cell, microglial cell, macrophage, monocyte, epithelial cell, endothelial cell, hepatocyte, spermatocyte, spermatid or spermatozoa.

[0049] Target cells may be transduced either *in vitro* after isolation from a human individual or may be transduced directly *in vivo*.

[0050] Viral vectors for use according to the invention are HIV vectors.

[0051] The term "derivable" is used in its normal sense as meaning a nucleotide sequence such as an LTR or a part thereof

which need not necessarily be obtained from a vector such as a retroviral vector but instead could be derived therefrom. By way of example, the sequence may be prepared synthetically or by use of recombinant DNA techniques.

[0052] Details on the genomic structure of some retroviruses may be found in the art. By way of example, details on HIV, and Mo-MLV may be found from the NCBI Genbank (Genome Accession Nos. AF033819 and AF033811, respectively). Details of HIV variants may also be found at <http://hiv-web.lanl.gov>. Details of ELAV variants may be found through <http://www.ncbi.nlm.nih.gov>.

[0053] The basic structure of a retrovirus genome is a 5' LTR and a 3' LTR, between or within which are located a packaging signal to enable the genome to be packaged, a primer binding site, integration sites to enable integration into a host cell genome and *gag*, *pol* and *env* genes encoding the packaging components - these are polypeptides required for the assembly of viral particles. More complex retroviruses have additional features, such as *rev* and RRE sequences in HIV, which enable the efficient export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell.

[0054] In the provirus, these genes are flanked at both ends by regions called long terminal repeats (LTRs). The LTRs are responsible for proviral integration, and transcription. LTRs also serve as enhancer-promoter sequences and can control the expression of the viral genes. Encapsidation of the retroviral RNAs occurs by virtue of a *psi* sequence, which it has been disclosed in respect of HIV, at least, is located at the 5' end of the viral genome.

[0055] The LTRs themselves are identical sequences that can be divided into three elements, which are called U3, R and U5. U3 is derived from the sequence unique to the 3' end of the RNA. R is derived from a sequence repeated at both ends of the RNA and U5 is derived from the sequence unique to the 5' end of the RNA. The sizes of the three elements can vary considerably among different retroviruses.

[0056] In a defective retroviral vector genome *gag*, *pol* and *env* may be absent or not functional. The R regions at both ends of the RNA are repeated sequences. U5 and U3 represent unique sequences at the 5' and 3' ends of the RNA genome respectively.

[0057] As discussed above, in a typical retroviral vector for use in gene therapy, at least part of one or more of the *gag*, *pol* and *env* protein coding regions essential for replication may be removed from the viral vector. This makes the retroviral vector replication-defective. The removed portions may even be replaced by a nucleotide sequence of interest (NOI), as in the present invention, to generate a virus capable of integrating its genome into a host genome but wherein the modified viral genome is unable to propagate itself due to a lack of structural proteins. When integrated in the host genome, expression of the NOI occurs - resulting in, for example, a therapeutic and/or a diagnostic effect. Thus, the transfer of an NOI into a site of interest is typically achieved by: integrating the NOI into the recombinant viral vector; packaging the modified viral vector into a virion coat; and allowing transduction of a site of interest - such as a targeted cell or a targeted cell population.

[0058] A minimal retroviral genome for use in the present invention may therefore comprise (5') R - U5 - one or more NOIs - U3-R (3'). However, the plasmid vector used to produce the retroviral genome within a host cell/packaging cell will also include transcriptional regulatory control sequences operably linked to the retroviral genome to direct transcription of the genome in a host cell/packaging cell. These regulatory sequences may be the natural sequences associated with the transcribed retroviral sequence, i.e. the 5' U3 region, or they may be a heterologous promoter such as another viral promoter, for example the CMV promoter.

[0059] Some retroviral genomes require additional sequences for efficient virus production. For example, in the case of HIV, *rev* and RRE sequence should be included. However, we have found that the requirement for *rev* and RRE can be reduced or eliminated by codon optimisation. As expression of the codon optimised *gag-pol* is REV independent, RRE can be removed from the *gag-pol* expression cassette, thus removing any potential for recombination with any RRE contained on the vector genome.

[0060] Once the retroviral vector NOIs sequences need to be expressed. In a retrovirus, the promoter is located in the 5' LTR U3 region of the provirus. In retroviral vectors, the promoter driving expression of a therapeutic gene may be the native retroviral promoter in the 5' U3 region, or an alternative promoter engineered into the vector. The alternative promoter may physically replace the 5' U3 promoter native to the retrovirus, or it may be incorporated at a different place within the vector genome such as between the LTRs.

[0061] Thus, the NOI will also be operably linked to a transcriptional regulatory control sequence to allow transcription of the first nucleotide sequence to occur in the target cell. The control sequence will typically be active in mammalian cells. The control sequence may, for example, be a viral promoter such as the natural viral promoter or a CMV promoter or it may be a mammalian promoter. It is particularly preferred to use a promoter that is preferentially active in a particular cell type or tissue type in which

the virus to be treated primarily infects. Thus, in one embodiment, a tissue-specific regulatory sequences may be used. The regulatory control sequences driving expression of the one or more first nucleotide sequences may be constitutive or regulated promoters.

[0062] The term "operably linked" denotes a relationship between a regulatory region (typically a promoter element, but may include an enhancer element) and the coding region of a gene, whereby the transcription of the coding region is under the control of the regulatory region.

[0063] As used herein, the term "enhancer" includes a DNA sequence which binds other protein components of the transcription initiation complex and thus facilitates the initiation of transcription directed by its associated promoter.

[0064] In one preferred embodiment of the present invention, the enhancer is an ischaemic like response element (ILRE).

[0065] The term "ischaemia like response element" - otherwise written as ILRE - includes an element that is responsive to or is active under conditions of ischaemia or conditions that are like ischaemia or are caused by ischaemia. By way of example, conditions that are like ischaemia or are caused by ischaemia include hypoxia and/or low glucose concentration(s).

[0066] The term "hypoxia" means a condition under which a particular organ or tissue receives an inadequate supply of oxygen.

[0067] Ischaemia can be an insufficient supply of blood to a specific organ or tissue. A consequence of decreased blood supply is an inadequate supply of oxygen to the organ or tissue (hypoxia). Prolonged hypoxia may result in injury-to the affected organ or tissue.

[0068] A preferred ILRE is a hypoxia response element (HRE).

[0069] In one preferred aspect of the present invention, there is hypoxia or ischaemia regulatable expression of the retroviral vector components. In this regard, hypoxia is a powerful regulator of gene expression in a wide range of different cell types and acts by the induction of the activity of hypoxia-inducible transcription factors such as hypoxia inducible factor-1 (HIF-1; 6), which bind to cognate DNA recognition sites, the hypoxia-responsive elements (HREs) on various gene promoters. Dachs *et al* (7) have used a multimeric form of the HRE from the mouse phosphoglycerate kinase-1 (PGK-1) gene (8) to control expression of both marker and therapeutic genes by human fibrosarcoma cells in response to hypoxia *in vitro* and within solid tumours *in vivo* (7 *ibid*).

[0070] Hypoxia response enhancer elements (HREEs) have also been found in association with a number of genes including the erythropoietin (EPO) gene (9; 10). Other HREEs have been isolated from regulatory regions of both the muscle glycolytic enzyme pyruvate kinase (PKM) gene (11), the human muscle-specific β -enolase gene (ENO3; 12) and the endothelin-1 (ET-1) gene (13).

[0071] Preferably the HRE for use according to the present invention is selected from, for example, the erythropoietin HRE element (HREE1), muscle pyruvate kinase (PKM), HRE element, phosphoglycerate kinase (PGK) HRE, B-enolase (enolase 3; ENO3) HRE element, endothelin-1 (ET-1)HRE element and metallothionein II (MTII) HRE element.

[0072] Preferably the ILRE is used in combination with a transcriptional regulatory element, such as a promoter, which transcriptional regulatory element is preferably active in one or more selected cell type(s), preferably being only active in one cell type.

[0073] As outlined above, this combination aspect is called a responsive element.

[0074] Preferably the responsive element comprises at least the ILRE as herein defined.

[0075] Non-limiting examples of such a responsive element are presented as OBHRE1 and XiaMac. Another non-limiting example includes the ILRE in use in conjunction with an MLV promoter and/or a tissue restricted ischaemic responsive promoter. These responsive elements are disclosed in WO99/15684.

[0076] Other examples of suitable tissue restricted promoters/enhancers are those which are highly active in tumour cells such as a promoter/enhancer from a *MUC1* gene, a *CEA* gene or a *5T4* antigen gene. The alpha-fetoprotein (AFP) promoter is also a tumour-specific promoter. One preferred promoter-enhancer combination is a human cytomegalovirus (hCMV) major immediate early (MIE) promoter/enhancer combination.

[0077] The term "promoter" is used in the normal sense of the art, e.g. an RNA polymerase binding site.

[0078] The promoter may be located in the retroviral 5' LTR to control the expression of a cDNA encoding an NOI, and/or gag-pol proteins.

[0079] Preferably the NOI and/or gag-pol proteins are capable of being expressed from the retrovirus genome such as from endogenous retroviral promoters in the long terminal repeat (LTR).

[0080] Preferably the NOI and/or gag-pol proteins are expressed from a heterologous promoter to which the heterologous gene or sequence, and/or codon optimised *gag-pol* sequence is operably linked.

[0081] Alternatively, the promoter may be an internal promoter.

[0082] Preferably the NOI is expressed from an internal promoter.

[0083] Vectors containing internal promoters have also been widely used to express multiple genes. An internal promoter makes it possible to exploit promoter/enhancer combinations other than those found in the viral LTR for driving gene expression. Multiple internal promoters can be included in a retroviral vector and it has proved possible to express at least three different cDNAs each from its own promoter (14). Internal ribosomal entry site (IRES) elements have also been used to allow translation of multiple coding regions from either a single mRNA or from fusion proteins that can then be expressed from an open reading frame.

[0084] The promoter may be constitutively efficient, or may be tissue or temporally restricted in their activity.

[0085] Preferably the promoter is a constitutive promoter such as CMV.

[0086] Preferably the promoters are tissue specific. That is, they are capable of driving transcription of a NOI or NOI(s) in one tissue while remaining largely "silent" in other tissue types.

[0087] The term "tissue specific" means a promoter which is not restricted in activity to a single tissue type but which nevertheless shows selectivity in that they may be active in one group of tissues and less active or silent in another group.

[0088] The level of expression of an NOI or NOIs under the control of a particular promoter may be modulated by manipulating the promoter region. For example, different domains within a promoter region may possess different gene regulatory activities. The roles of these different regions are typically assessed using vector constructs having different variants of the promoter with specific regions deleted (that is, deletion analysis). This approach may be used to identify, for example, the smallest region capable of conferring tissue specificity or the smallest region conferring hypoxia sensitivity.

[0089] A number of tissue specific promoters, described above, may be particularly advantageous in practising the present invention. In most instances, these promoters may be isolated as convenient restriction digestion fragments suitable for cloning in a selected vector. Alternatively, promoter fragments may be isolated using the polymerase chain reaction. Cloning of the amplified fragments may be facilitated by incorporating restriction sites at the 5' end of the primers.

[0090] The NOI or NOIs may be under the expression control of an expression regulatory element, such as a promoter and enhancer.

[0091] Preferably the ischaemic responsive promoter is a tissue restricted ischaemic responsive promoter.

[0092] Preferably the tissue restricted ischaemic responsive promoter is a macrophage specific promoter restricted by repression.

[0093] Preferably the tissue restricted ischaemic responsive promoter is an endothelium specific promoter.

[0094] Preferably the regulated retroviral vector of the present invention is an ILRE regulated retroviral vector.

[0095] Preferably the regulated retroviral vector for use according to the present invention is an ILRE regulated lentiviral vector.

[0096] Preferably the regulated retroviral vector for use according to the present invention is an

autoregulated hypoxia responsive lentiviral vector.

[0097] Preferably the regulated retroviral vector for use according to the present invention is regulated by glucose concentration.

[0098] For example, the glucose-regulated proteins (grp's) such as grp78 and grp94 are highly conserved proteins known to be induced by glucose deprivation (15). The grp 78 gene is expressed at low levels in most normal healthy tissues under the influence of basal level promoter elements but has at least two critical "stress inducible regulatory elements" upstream of the TATA element (15 *ibid*; 16). Attachment to a truncated 632 base pair sequence of the 5' end of the grp78 promoter confers high inducibility to glucose deprivation on reporter genes *in vitro* (16 *ibid*). Furthermore, this promoter sequence in retroviral vectors was capable of driving a high level expression of a reporter gene in tumour cells in murine fibrosarcomas, particularly in central relatively ischaemic/fibrotic sites (16 *ibid*).

[0099] The regulated retroviral vector for use according to the present invention is a self-inactivating (SIN) vector.

[0100] By way of example, self-inactivating retroviral vectors have been constructed by deleting the transcriptional enhancers or the enhancers and promoter in the U3 region of the 3' LTR. After a round of vector reverse transcription and integration, these changes are copied into both the 5' and the 3' LTRs producing a transcriptionally inactive provirus (17; 18; 19; 20). However, any promoter(s) internal to the LTRs in such vectors will still be transcriptionally active. This strategy has been employed to eliminate effects of the enhancers and promoters in the viral LTRs on transcription from internally placed genes. Such effects include increased transcription (21) or suppression of transcription (22). This strategy can also be used to eliminate downstream transcription from the 3' LTR into genomic DNA (23). This is of particular concern in human gene therapy where it is of critical importance to prevent the adventitious activation of an endogenous oncogene.

[0101] As discussed above, replication-defective retroviral vectors are typically propagated, for example to prepare suitable titres of the retroviral vector for subsequent transduction, by using a combination of a packaging or helper cell line and the recombinant vector. That is to say, that the three packaging proteins can be provided *in trans*.

[0102] In general a "packaging cell line" contains one or more of the retroviral *gag*, *pol* and *env* genes. In the present invention it contains codon optimised *gag-pol* genes, and an *env* gene. The packaging cell line produces the proteins required for packaging retroviral DNA but it cannot bring about encapsidation. Conventionally this has been achieved through lack of a *psi* region. However, when a recombinant vector carrying an NOI and a *psi* region is introduced into the packaging cell line, the helper proteins can package the *psi*-positive recombinant vector to produce the recombinant virus stock. This virus stock can be used to transduce cells to introduce the NOI into the genome of the target cells. Conventionally a *psi* packaging signal, called *psi* plus, has been used that contains additional sequences spanning from upstream of the splice donor to downstream of the *gag* start codon (24) since this has been shown to increase viral titres.

[0103] The recombinant virus whose genome lacks all genes required to make viral proteins can transduce only once and cannot propagate. These viral vectors which are only capable of a single round of transduction of target cells are known as replication defective vectors. Hence, the NOI is introduced into the host/target cell genome without the generation of potentially harmful retrovirus. A summary of the available packaging lines is presented in Coffin *et al.*, 1997 (*ibid*).

[0104] The retroviral packaging cell line is preferably in the form of a transiently transfected cell line. Transient transfections may advantageously be used to measure levels of vector production when vectors are being developed. In this regard, transient transfection avoids the longer time required to generate stable vector-producing cell lines and may also be used if the vector or retroviral packaging components are toxic to cells. Components typically used to generate retroviral vectors include a plasmid encoding the gag-pol proteins, a plasmid encoding the env protein and a plasmid containing an NOI. Vector production involves transient transfection of one or more of these components into cells containing the other required components. If the vector encodes toxic genes or genes that interfere with the replication of the host cell, such as inhibitors of the cell cycle or genes that induce apoptosis, it may be difficult to generate stable vector-producing cell lines, but transient transfection can be used to produce the vector before the cells die. Also, cell lines have been developed using transient transfection that produce vector titre levels that are comparable to the levels obtained from stable vector-producing cell lines (25).

[0105] Producer cells/packaging cells can be of any suitable cell type. Producer cells are generally mammalian cells but can be, for example, insect cells. A producer cell may be a packaging cell containing the virus structural genes, normally integrated into its genome into which the regulated retroviral vectors of the present invention are introduced. Alternatively the producer cell may be transfected with nucleic acid sequences encoding structural components, such as codon optimised *gag-pol* and *env* on one or

more vectors such as plasmids, adenovirus vectors, herpes viral vectors or any method known to deliver functional DNA into target cells. The vectors for use according to the present invention are then introduced into the packaging cell by the methods of the present invention.

[0106] As used herein, the term "producer cell" or "vector producing cell" refers to a cell which contains all the elements necessary for production of regulated retroviral vector particles and regulated retroviral delivery systems.

[0107] Preferably, the producer cell is obtainable from a stable producer cell line. Preferably, the producer cell is obtainable from a derived stable producer cell line.

[0108] Preferably, the producer cell is obtainable from a derived producer cell line. As used herein, the term "derived producer cell line" is a transduced producer cell line which has been screened and selected for high expression of a marker gene. Such cell lines contain retroviral insertions in integration sites that support high level expression from the retroviral genome. The term "derived producer cell line" is used interchangeably with the term "derived stable producer cell line" and the term "stable producer cell line".

[0109] The derived producer cell line is an HIV producer cell line.

[0110] Preferably the envelope protein sequences, and nucleocapsid sequences are all stably integrated in the producer and/or packaging cell. However, one or more of these sequences could also exist in episomal form and gene expression could occur from the episome.

[0111] As used herein, the term "packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant virus which are lacking in a recombinant viral vector. Typically, such packaging cells contain one or more vectors which are capable of expressing viral structural proteins (such as codon optimised *gag-pol* and *env*) but they do not contain a packaging signal.

[0112] The term "packaging signal" which is referred to interchangeably as "packaging sequence" or "*psi*" is used in reference to the non-coding, *cis*-acting sequence required for encapsidation of retroviral RNA strands during viral particle formation. In HIV-1, this sequence has been mapped to loci extending from upstream of the major splice donor site (SD) to at least the *gag* start codon.

[0113] Packaging cell lines suitable for use with the above-described vector constructs may be readily prepared (see also WO 92/05266), and utilised to create producer cell lines for the production of retroviral vector particles. As already mentioned, a summary of the available packaging lines is presented in "Retroviruses" (1997 Cold Spring Harbour Laboratory Press Eds: JM Coffin, SM Hughes, HE Varmus pp 449).

[0114] Also as discussed above, simple packaging cell lines, comprising a provirus in which the packaging signal has been deleted, have been found to lead to the rapid production of undesirable replication competent viruses through recombination. In order to improve safety, second generation cell lines have been produced wherein the 3'LTR of the provirus is deleted. In such cells, two recombinations would be necessary to produce a wild type virus. A further improvement involves the introduction of the *gag-pol* genes and the *env* gene on separate constructs so-called third generation packaging cell lines. These constructs are introduced sequentially to prevent recombination during transfection (26; 27).

[0115] Preferably, the packaging cell lines are second generation packaging cell lines.

[0116] Preferably, the packaging cell lines are third generation packaging cell lines.

[0117] In these split-construct, third generation cell lines, a further reduction in recombination may be achieved by "codon wobbling". This technique, based on the redundancy of the genetic code, aims to reduce homology between the separate constructs, for example between the regions of overlap in the *gag-pol* and *env* open reading frames.

[0118] The packaging cell lines are useful for providing the gene products necessary to encapsidate and provide a membrane protein for a high titre regulated retrovirus vector and regulated nucleic gene delivery vehicle production. When regulated retrovirus sequences are introduced into the packaging cell lines, such sequences are encapsidated with the nucleocapsid (*gag-pol*) proteins and these units then bud through the cell membrane to become surrounded in cell membrane and to contain the envelope protein produced in the packaging cell line. These infectious regulated retroviruses are useful as infectious units *per se*.

or as gene delivery vectors.

[0119] The packaging cell may be a cell cultured *in vitro* such as a tissue culture cell line. Suitable cell lines include but are not limited to mammalian cells such as murine fibroblast derived cell lines or human cell lines. Preferably the packaging cell line is a human cell line, such as for example: HEK293, 293-T, TE671, HT1080.

[0120] Alternatively, the packaging cell may be a cell derived from the individual to be treated such as a monocyte, macrophage, blood cell or fibroblast. The cell may be isolated from an individual and the packaging and vector components administered *ex vivo* followed by re-administration of the autologous packaging cells.

[0121] It is highly desirable to use high-titre virus preparations in both experimental and practical applications. Techniques for increasing viral titre include using a *psi* plus packaging signal as discussed above and concentration of viral stocks. In addition, the use of different envelope proteins, such as the G protein from vesicular-stomatitis virus has improved titres following concentration to 10^9 per ml (28). However, typically the envelope protein will be chosen such that the viral particle will preferentially infect cells that are infected with the virus which it desired to treat. For example where an HIV vector is being used to treat HIV infection, the *env* protein used will be the HIV *env* protein.

[0122] The process of producing a retroviral vector in which the envelope protein is not the native envelope of the retrovirus is known as "pseudotyping". Certain envelope proteins, such as MLV envelope protein and vesicular stomatitis virus G (VSV-G) protein, pseudotype retroviruses very well. Pseudotyping is not a new phenomenon and examples may be found in WO-A-98/05759, WO-A-98/05754, WO-A-97/17457, WO-A-96/09400, WO-A-91/00047 and (29).

[0123] As used herein, the term "high titre" means an effective amount of a retroviral vector or particle which is capable of transducing a target site such as a cell.

[0124] As used herein, the term "effective amount" means an amount of a regulated retroviral or lentiviral vector or vector particle which is sufficient to induce expression of an NOI at a target site.

[0125] Preferably the titre is from at least 10^6 retrovirus particles per ml, such as from 10^6 to 10^7 per ml, more preferably at least 10^7 retrovirus particles per ml.

[0126] In accordance with the present invention, it is possible to manipulate the viral genome or the regulated retroviral vector nucleotide sequence, so that viral genes are replaced or supplemented with one or more NOIs which may be heterologous NOIs.

[0127] The term "heterologous" refers to a nucleic acid sequence or protein sequence linked to a nucleic acid or protein sequence which it is not naturally linked.

[0128] With the present invention, the term NOI (i.e. nucleotide sequence of interest) includes any suitable nucleotide sequence, which need not necessarily be a complete naturally occurring DNA sequence. Thus, the DNA sequence can be, for example, a synthetic DNA sequence, a recombinant DNA sequence (i.e. prepared by use of recombinant DNA techniques), a cDNA sequence or a partial genomic DNA sequence, including combinations thereof. The DNA sequence need not be a coding region. If it is a coding region, it need not be an entire coding region. In addition, the DNA sequence can be in a sense orientation or in an anti-sense orientation. Preferably, it is in a sense orientation. Preferably, the DNA is or comprises cDNA.

[0129] The NOI(s) may be any one or more of selection gene(s), marker gene(s) and therapeutic gene(s).

[0130] As used herein, the term "selection gene" refers to the use of a NOI which encodes a selectable marker which may have an enzymatic activity that confers resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed.

[0131] Many different selectable markers have been used successfully in retroviral vectors. These are reviewed in "retroviruses" (1997 Cold Spring Harbour Laboratory Press Eds: JM Coffin, SM Hughes, HE Varmus pp 444) and include, but are not limited to, the bacterial neomycin (neo) and hygromycin phosphotransferase genes which confer resistance to G418 and hygromycin respectively; a mutant mouse dihydrofolate reductase gene which confers resistance to methotrexate; the bacterial *gpt* gene which allows cells to grow in medium containing mycophenolic acid, xanthine and aminopterin; the bacterial *hisD* gene which allows cells to grow in medium without histidine but containing histidinol; the multidrug resistance gene (*mdr*) which confers resistance to a variety of drugs; and the bacterial genes which confer resistance to puromycin or phleomycin. All of these markers are dominant selectable and allow chemical selection of most cells expressing these genes. Other selectable markers are not

dominant in that their use must be in conjunction with a cell line that lacks the relevant enzyme activity. Examples of non-dominant selectable markers include the thymidine kinase (*tk*) gene which is used in conjunction with *tk* cell lines.

[0132] Particularly preferred markers are blasticidin and neomycin, optionally operably linked to a thymidine kinase coding sequence typically under the transcriptional control of a strong viral promoter such as the SV40 promoter.

[0133] In accordance with the present invention, suitable NOI sequences include those that are of therapeutic and/or diagnostic application such as, but are not limited to: sequences encoding cytokines, chemokines, hormones, antibodies, engineered immunoglobulin-like molecules, a single chain antibody, fusion proteins, enzymes, immune co-stimulatory molecules, immunomodulatory molecules, anti-sense RNA, a transdominant negative mutant of a target protein, a toxin, a conditional toxin, an antigen, a tumour suppressor protein and growth factors, membrane proteins, vasoactive proteins and peptides, anti-viral proteins and ribozymes, and derivatives thereof (such as with an associated reporter group). When included, such coding sequences may be typically operatively linked to a suitable promoter, which may be a promoter driving expression of a ribozyme(s), or a different promoter or promoters, such as in one or more specific cell types.

[0134] Suitable NOIs for use in the invention in the treatment or prophylaxis of cancer include NOIs encoding proteins which: destroy the target cell (for example a ribosomal toxin), act as: tumour suppressors (such as wild-type p53); activators of anti-tumour immune mechanisms (such as cytokines, co-stimulatory molecules and immunoglobulins); inhibitors of angiogenesis; or which provide enhanced drug sensitivity (such as pro-drug activation enzymes); indirectly stimulate destruction of target cell by natural effector cells (for example, strong antigen to stimulate the immune system or convert a precursor substance to a toxic substance which destroys the target cell (for example a prodrug activating enzyme).

[0135] Examples of prodrugs include but are not limited to etoposide phosphate (used with alkaline phosphatase; 5-fluorocytosine (with cytosine deaminase); Doxorubin-N-p-hydroxyphenoxyacetamide (with Penicillin-V-Amidase); Para-N-bis (2-chloroethyl)aminobenzoyl glutamate (with Carboxypeptidase G2); Cephalosporin nitrogen mustard carbamates (with B-lactamase); SR4233 (with p450 reductase); Ganciclovir (with HSV thymidine kinase); mustard pro-drugs with nitroreductase and cyclophosphamide or ifosfamide (with cytochrome p450).

[0136] Suitable NOIs for use in the treatment or prevention of ischaemic heart disease include NOIs encoding plasminogen activators. Suitable NOIs for the treatment or prevention of rheumatoid arthritis or cerebral malaria include genes encoding anti-inflammatory proteins, antibodies directed against tumour necrosis factor (TNF) alpha, and anti-adhesion molecules (such as antibody molecules or receptors specific for adhesion molecules).

[0137] The expression products encoded by the NOIs may be proteins which are secreted from the cell. Alternatively the NOI expression products are not secreted and are active within the cell. In either event, it is preferred for the NOI expression product to demonstrate a bystander effect or a distant bystander effect; that is the production of the expression product in one cell leading to the killing of additional, related cells, either neighbouring or distant (e.g. metastatic), which possess a common phenotype. Encoded proteins could also destroy bystander tumour cells (for example with secreted antitumour antibody-ribosomal toxin fusion protein), indirectly stimulated destruction of bystander tumour cells (for example cytokines to stimulate the immune system or procoagulant proteins causing local vascular occlusion) or convert a precursor substance to a toxic substance which destroys bystander tumour cells (eg an enzyme which activates a prodrug to a diffusible drug). Also, the delivery of NOI(s) encoding antisense transcripts or ribozymes which interfere with expression of cellular genes for tumour persistence (for example against aberrant *myc* transcripts in Burkitts lymphoma or against *bcr-abl* transcripts in chronic myeloid leukemia. The use of combinations of such NOIs is also envisaged.

[0138] The NOI or NOIs for use in the present invention may also comprise one or more cytokine-encoding NOIs. Suitable cytokines and growth factors include but are not limited to: ApoE, Apo-SAA, BDNF, Cardiotrophin-1, EGF, ENA-78, Eotaxin, Eotaxin-2, Exodus-2, FGF-acidic, FGF-basic, fibroblast growth factor-10 (30), FLT3 ligand, Fractalkine (CX3C), GDNF, G-CSF, GM-CSF, GF-β1, insulin, IFN-γ, IGF-I, IGF-II, IL-1α, IL-1β, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8 (72 a.a.), IL-8 (77 a.a.), IL-9, IL-10, IL-11, IL-12, IL-13, IL-15, IL-16, IL-17, IL-18 (IGIF), Inhibin α, Inhibin β, IP-10, keratinocyte growth factor-2 (KGF-2), KGF, Leptin, LIF, Lymphotoxin, Mullerian inhibitory substance, monocyte colony inhibitory factor, monocyte attractant protein (30 *ibid*), M-CSF, MDC (67 a.a.), MDC (69 a.a.), MCP-1 (MCAF), MCP-2, MCP-3, MCP-4, MDC (67 a.a.), MDC (69 a.a.), MIG, MIP-1α, MIP-1β, MIP-3α, MIP-3β, MIP-4, myeloid progenitor inhibitor factor-1 (MPIF-1), NAP-2, Neurturin, Nerve growth factor, β-NGF, NT-3, NT-4, Oncostatin M, PDGF-AA, PDGF-AB, PDGF-BB, PF-4, RANTES, SDF1α, SDF1β, SCF, SCGF, stem cell factor (SCF), TARC, TGF-α, TGF-β, TGF-β2, TGF-β3, tumour necrosis factor (TNF), TNF-α, TNF-β, TNIL-1, TPO, VEGF, GCP-2, GRO/MGSA, GRO-β, GRO-γ, HCC1, 1-309.

[0139] The NOI or NOIs may be under the expression control of an expression regulatory element, such as a promoter and/or a

promoter enhancer as known as "responsive elements".

[0140] When the regulated retroviral vector particles are used to transfer NOIs into cells which they transduce, such vector particles also designated "viral delivery systems" or "retroviral delivery systems". Viral vectors, including retroviral vectors, have been used to transfer NOIs efficiently by exploiting the viral transduction process. NOIs cloned into the retroviral genome can be delivered efficiently to cells susceptible to transduction by a retrovirus. Through other genetic manipulations, the replicative capacity of the retroviral genome can be destroyed. The vectors introduce new genetic material into a cell but are unable to replicate.

[0141] The regulated retroviral vector for use according to the present invention can be delivered by viral or non-viral techniques. Non-viral delivery systems include but are not limited to DNA transfection methods. Here, transfection includes a process using a non-viral vector to deliver a gene to a target mammalian cell.

[0142] Typical transfection methods include electroporation, DNA biolistics, lipid-mediated transfection, compacted DNA-mediated transfection, liposomes, immunoliposomes, lipofectin, cationic agent-mediated, cationic facial amphiphiles (CFAs) (31), multivalent cations such as spermine, cationic lipids or polylysine, 1, 2-bis (oleoyloxy)-3-(trimethylammonio) propane (DOTAP)-cholesterol complexes (32) and combinations thereof.

[0143] Viral delivery systems include but are not limited to adenovirus vector, an adeno-associated viral (AAV) vector, a herpes viral vector, a retroviral vector, a lentiviral vector, or a baculoviral vector. These viral delivery systems may be configured as a split-intron vector. A split intron vector is described in WO 99/15683.

[0144] Other examples of vectors include *ex vivo* delivery systems, which include but are not limited to DNA transfection methods such as electroporation, DNA biolistics, lipid-mediated transfection, compacted DNA-mediated transfection.

[0145] The vector may be a plasmid DNA vector. Alternatively, the vector may be a recombinant viral vector. Suitable recombinant viral vectors include adenovirus vectors, adeno-associated viral (AAV) vectors, Herpes-virus vectors, or retroviral vectors, lentiviral vectors or a combination of adenoviral and lentiviral vectors. In the case of viral vectors, gene delivery is mediated by viral infection of a target cell.

[0146] If the features of adenoviruses are combined with the genetic stability of retro/lentiviruses then essentially the adenovirus can be used to transduce target cells to become transient retroviral producer cells that could stably infect neighbouring cells.

[0147] Also disclosed is a pharmaceutical composition for treating an individual by gene therapy, wherein the composition comprises a therapeutically effective amount of a regulated retroviral vector as described. The pharmaceutical composition may be for human or animal usage. Typically, a physician will determine the actual dosage which will be most suitable for an individual subject and it will vary with the age, weight and response of the particular patient.

[0148] The composition may optionally comprise a pharmaceutically acceptable carrier, diluent, excipient or adjuvant. The choice of pharmaceutical carrier, excipient or diluent can be selected with regard to the intended route of administration and standard pharmaceutical practice. The pharmaceutical compositions may comprise as - or in addition to - the carrier, excipient or diluent any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s), solubilising agent(s), and other carrier agents that may aid or increase the viral entry into the target site (such as for example a lipid delivery system).

[0149] Where appropriate, the pharmaceutical compositions can be administered by any one or more of: minipumps; inhalation, in the form of a suppository or pessary, topically in the form of a lotion, solution, cream, ointment or dusting powder, by use of a skin patch, orally in the form of tablets containing excipients such as starch or lactose, or in capsules or ovules either alone or in admixture with excipients, or in the form of elixirs, solutions or suspensions containing flavouring or colouring agents, or they can be injected parenterally, for example intracavernosally, intravenously, intramuscularly or subcutaneously. For parenteral administration, the compositions may be best used in the form of a sterile aqueous solution which may contain other substances, for example enough salts or monosaccharides to make the solution isotonic with blood. For buccal or sublingual administration the compositions may be administered in the form of tablets or lozenges which can be formulated in a conventional manner.

[0150] The present invention is believed to have a wide therapeutic applicability - depending on *inter alia* the selection of the one or more NOIs.

[0151] For example, the present invention may be useful in the treatment of the disorders listed in WO-A-98/05635. For ease of

reference, part of that list is now provided: cancer, inflammation or inflammatory disease, dermatological disorders, fever, cardiovascular effects, haemorrhage, coagulation and acute phase response, cachexia, anorexia, acute infection, HIV infection, shock states, graft-versus-host reactions, autoimmune disease, reperfusion injury, meningitis, migraine and aspirin-dependent anti-thrombosis; tumour growth, invasion and spread, angiogenesis, metastases, malignant, ascites and malignant pleural effusion; cerebral ischaemia, ischaemic heart disease, osteoarthritis, rheumatoid arthritis, osteoporosis, asthma, multiple sclerosis, neurodegeneration, Alzheimer's disease, atherosclerosis, stroke, vasculitis, Crohn's disease and ulcerative colitis; periodontitis, gingivitis; psoriasis, atopic dermatitis, chronic ulcers, epidermolysis bullosa; corneal ulceration, retinopathy and surgical wound healing; rhinitis, allergic conjunctivitis, eczema, anaphylaxis; restenosis, congestive heart failure, endometriosis, atherosclerosis or endosclerosis.

[0152] In addition, or in the alternative, the present invention may be useful in the treatment of disorders listed in WO-A-98/07859. For ease of reference, part of that list is now provided: cytokine and cell proliferation/differentiation activity; immunosuppressant or immunostimulant activity (e.g. for treating immune deficiency, including infection with human immune deficiency virus; regulation of lymphocyte growth; treating cancer and many autoimmune diseases, and to prevent transplant rejection or induce tumour immunity); regulation of haematopoiesis, e.g. treatment of myeloid or lymphoid diseases; promoting growth of bone, cartilage, tendon, ligament and nerve tissue, e.g. for healing wounds, treatment of burns, ulcers and periodontal disease and neurodegeneration; inhibition or activation of follicle-stimulating hormone (modulation of fertility); chemotactic/chemokinetic activity (e.g. for mobilising specific cell types to sites of injury or infection); haemostatic and thrombolytic activity (e.g. for treating haemophilia and stroke); antiinflammatory activity (for treating e.g. septic shock or Crohn's disease); as antimicrobials; modulators of e.g. metabolism or behaviour; as analgesics; treating specific deficiency disorders; in treatment of e.g. psoriasis, in human or veterinary medicine.

[0153] In addition, or in the alternative, the present invention may be useful in the treatment of disorders listed in WO-A-98/09985. For ease of reference, part of that list is now provided: macrophage inhibitory and/or T cell inhibitory activity and thus, anti-inflammatory activity; anti-immune activity, i.e. inhibitory effects against a cellular and/or humoral immune response, including a response not associated with inflammation; inhibit the ability of macrophages and T cells to adhere to extracellular matrix components and fibronectin, as well as up-regulated fas receptor expression in T cells; inhibit unwanted immune reaction and inflammation including arthritis, including rheumatoid arthritis, inflammation associated with hypersensitivity, allergic reactions, asthma, systemic lupus erythematosus, collagen diseases and other autoimmune diseases, inflammation associated with atherosclerosis, arteriosclerosis, atherosclerotic heart disease, reperfusion injury, cardiac arrest, myocardial infarction, vascular inflammatory disorders, respiratory distress syndrome or other cardiopulmonary diseases, inflammation associated with peptic ulcer, ulcerative colitis and other diseases of the gastrointestinal tract, hepatic fibrosis, liver cirrhosis or other hepatic diseases, thyroiditis or other glandular diseases, glomerulonephritis or other renal and urologic diseases, otitis or other oto-rhino-laryngological diseases, dermatitis or other dermal diseases, periodontal diseases or other dental diseases, orchitis or epididymo-orchitis, infertility, orchidial trauma or other immune-related testicular diseases, placental dysfunction, placental insufficiency, habitual abortion, eclampsia, pre-eclampsia and other immune and/or inflammatory-related gynaecological diseases, posterior uveitis, intermediate uveitis, anterior uveitis, conjunctivitis, chorioretinitis, uveoretinitis, optic neuritis, intraocular inflammation, e.g. retinitis or cystoid macular oedema, sympathetic ophthalmia, scleritis, retinitis pigmentosa, immune and inflammatory components of degenerative fundus disease, inflammatory components of ocular trauma, ocular inflammation caused by infection, proliferative vitreo-retinopathies, acute ischaemic optic neuropathy, excessive scarring, e.g. following glaucoma filtration operation, immune and/or inflammation reaction against ocular implants and other immune and inflammatory-related ophthalmic diseases, inflammation associated with autoimmune diseases or conditions or disorders where, both in the central nervous system (CNS) or in any other organ, immune and/or inflammation suppression would be beneficial, Parkinson's disease, complication and/or side effects from treatment of Parkinson's disease, AIDS-related dementia complex HIV-related encephalopathy, Devic's disease, Sydenham chorea, Alzheimer's disease and other degenerative diseases, conditions or disorders of the CNS, inflammatory components of strokes, post-polio syndrome, immune and inflammatory components of psychiatric disorders, myelitis, encephalitis, subacute sclerosing pan-encephalitis, encephalomyelitis, acute neuropathy, subacute neuropathy, chronic neuropathy, Guillain-Barre syndrome, Sydenham chorea, myasthenia gravis, pseudo-tumour cerebri, Down's Syndrome, Huntington's disease, amyotrophic lateral sclerosis, inflammatory components of CNS compression or CNS trauma or infections of the CNS, inflammatory components of muscular atrophies and dystrophies, and immune and inflammatory related diseases, conditions or disorders of the central and peripheral nervous systems, post-traumatic inflammation, septic shock, infectious diseases, inflammatory complications or side effects of surgery, bone marrow transplantation or other transplantation complications and/or side effects, inflammatory and/or immune complications and side effects of gene therapy, e.g. due to infection with a viral carrier, or inflammation associated with AIDS, to suppress or inhibit a humoral and/or cellular immune response, to treat or ameliorate monocyte or leukocyte proliferative diseases, e.g. leukaemia, by reducing the amount of monocytes or lymphocytes, for the prevention and/or treatment of graft rejection in cases of transplantation of natural or artificial cells, tissue and organs such as cornea, bone marrow, organs, lenses, pacemakers, natural or artificial skin tissue.

[0154] The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention. The Examples refer to the Figures. In the Figures:

Description of the Figures

[0155]

Figure 1 shows schematically how to create a suitable 3' LTR by PCR;

Figure 2 shows the codon usage table for wild type HIV *gag-pol* of strain HXB2 (accession number: K03455);

Figure 3a shows the codon usage table of the codon optimised sequence designated gagpol-SYNgp. Figure 3b shows a comparative codon usage table;

Figure 4 shows the codon usage table of the wild type HIV *env* called env-mn;

Figure 5 shows the codon usage table of the codon optimised sequence of HIV *env* designated SYNgp160mn;

Figure 6 shows two plasmid constructs for use in the invention;

Figure 7 shows the principle behind two systems for producing retroviral vector particles;

Figure 8 shows a sequence comparison between the wild type HIV *gag-pol* sequence (pGP-RRE3) and the codon optimised *gag-pol* sequence (pSYNGP);

Figure 9 shows a sequence comparison between the wild type EIAV *gag-pol* sequence (WT) and the codon optimised *gag-pol* sequence (CO);

Figure 10 shows Rev independence of protein expression particle formation;

Figure 11 shows translation rates of wild-type (WT) and codon optimised *gag-pol*;

Figure 12 shows *gag-pol* mRNA levels in total and cytoplasmic fractions;

Figure 13 shows the effect of insertion of WT *gag* downstream of the codon optimised gene on RNA and protein levels;

Figure 14 shows the plasmids used to study the effect of HIV-1 *gag* on the codon optimised gene;

Figure 15 shows the effect on cytoplasmic RNA of insertion of HIV-1 *gag* upstream of the codon optimised gene;

Figure 16 shows the effect of Leptomycin B (LMB) on protein production;

Figure 17 shows the cytoplasmic RNA levels of the vector genomes;

Figure 18 shows transduction efficiency at MOI 1;

Figure 19 shows a schematic representation of pGP-RRE3;

Figure 20 shows a schematic representation of pSYNGP;

Figure 21 shows vector titres generated with different *gag-pol* constructs;

Figure 22 shows vector titres from the Rev/RRE (-) and (+) genomes;

Figure 23 shows vector titres from the pHS series of vector genomes;

Figure 24 shows vector titres for the pHS series of vector genomes in the presence or absence of Rev/RRE;

Figure 25 shows an analysis of *gag-pol* constructs;

Figure 26 shows a Western blot of 293T extracts;

Figure 27 is a schematic representation of pESYNGP;

Figure 28 is a schematic representation of LpESYNGP;

Figure 29 is a schematic representation of LpESYNGPRRE;

Figure 30 is a schematic representation of pESYNGPRRE;

Figure 31 is a schematic representation of pONY4.0Z;

Figure 32 is a schematic representation of pONY8.0Z;

Figure 33 is a schematic representation of pONY8.1 Z;

Figure 34 is a schematic representation of pONY3.1;

Figure 35 is a schematic representation of pCIneoERev;

Figure 36 is a schematic representation of pESYNREV;

Figures 37 and 38 show the effect of different vector constructs on viral vector titres;

Figures 39 and 40 show the effect of different vector constructs on RT activity;

Figure 41 shows the effect of the 5' leader sequence on viral vector titre;

Figure 42 shows viral vector titres when using pONY8.1Z;

Figure 43 shows a comparison between the sequences of pONY3.1 and codon optimised pONY3.2OPTI in the first 372 nucleotides of *gag*;

Figure 44 is a schematic representation of pIRES1hygESYNGP;

Figure 45 and 46 show the results of experiments to confirm that codon optimised *gag-pol* can be used in the production of packaging and producer cell lines;

Figures 47 and 48 show the results of experiments which confirm that RNA from codon optimised *gag-pol* is packaged less efficiently than that from the wild-type gene;

Figure 49 shows the results of an experiment which confirms that expression from pESYNGP and pESDSYNGP are similar;

Figure 50 is a schematic representation of pESDSYNGP; and

Figure 51 shows the results of an experiment which confirms that the efficiency of encapsidating *gag-pol* RNA in PEV-17 cells and B-241 cells is similar.

[0156] In more detail, Figure 8 shows a sequence comparison between the wild type HIV *gag-pol* sequence (pGP-RRE3) and the codon optimised *gag-pol* sequence (pSYNGP) wherein the upper sequence represents pSYNGP and the lower sequence represents pGP-RRE3.

[0157] Figure 10 shows Rev independence of protein expression particle formation. 5µg of the *gag-pol* expression plasmids were transfected into 293T cells in the presence or absence of Rev (pCMV-Rev, 1µg) and protein levels were determined 48 hours post transfection in culture supernatants (A) and cell lysates (B). HIV-1 positive human serum was used to detect the *gag-pol* proteins. The blots were re-probed with an anti-actin antibody, as an internal control (C). The protein marker (New England Biolabs) sizes (in kDa) are shown on the side of the gel. Lanes: 1. Mock transfected 293T cells, 2. pGP-RRE3, 3. pGP-RRE3 + pCMV-Rev, 4. pSYNGP, 5. pSYNGP + pCMV-Rev, 6. pSYNGP-RRE, 7. pSYNGP-RRE + pCMV-Rev, 8. pSYNGP-ERR, 9. pSYNGP-ERR + pCMV-Rev.

[0158] Figure 11 shows translation rates of WT and codon optimised *gag-pol*. 293T cells were transfected with 2µg pGP-RRE3 (+/- 1µg pCMV-Rev) or 2µg pSYNGP. Protein samples from culture supernatants (A) and cell extracts (B) were analysed by Western blotting 12, 25, 37 and 48 hours post-transfection. HIV-1 positive human serum was used to detect *gag-pol* proteins (A, B) and an anti-actin antibody was used as an internal control (C). The protein marker sizes are shown on the side of the gel (in kD). A Phosphorimager was used for quantification of the results. Lanes: 1. pGP-RRE3 12h, 2. pGP-RRE3 25h, 3. pGP-RRE3 37h, 4. pGP-RRE3 48h, 5. pGP-RRE3 + pCMV-Rev 12h, 6. pGP-RRE3 + pCMV-Rev 25h, 7. pGP-RRE3 + pCMV-Rev 37h, 8.

pGP-RRE3 + pCMV-Rev 48h, 9. pSYNGP 12h, 10. pSYNGP 25h, 11. pSYNGP 37h, 12. pSYNGP 48h, 13. Mock transfected 293T cells.

[0159] Figure 12 shows *gag-pol* mRNA levels in total and cytoplasmic fractions. Total and cytoplasmic RNA was extracted from 293T cells 36 hours after transfection with 5µg of the *gag-pol* expression plasmid (+/- 1µg pCMV-Rev) and mRNA levels were estimated by Northern blot analysis. A probe complementary to nt 1222-1503 of both the wild type and codon optimised gene was used. Panel A shows the band corresponding to the HIV-1 *gag-pol*. The sizes of the mRNAs are 4.4 kb for the codon optimised and 6 kb for the wild type gene. Panel B shows the band corresponding to human ubiquitin (internal control for normalisation of results). Quantification was performed using a Phosphorimager. Lane numbering: c indicates cytoplasmic fraction and t indicates total RNA fraction. Lanes: 1. pGP-RRE3, 2. pGP-RRE3 + pCMV-Rev, 3. pSYNGP, 4. pSYNGP + pCMV-Rev, 5. pSYNGP-RRE, 6. pSYNGP-RRE + pCMV-Rev, 7. Mock transfected 293T cells, 8. pGP-RRE3 + pCMV-Rev, 9. Mock transfected 293T cells, 10. pSYNGP.

[0160] Figure 13 shows the effect of insertion of WT *gag* downstream of the codon optimised gene on RNA and protein levels. The wt *gag* sequence was inserted downstream of the codon optimised gene in both orientations (*NotI* site), resulting in plasmids pSYN6 (correct orientation, see Figure 14) and pSYN7 (reverse orientation, see Figure 14). The gene encoding for β-galactosidase (LacZ) was also inserted in the same site and the correct orientation (plasmid pSYN8, see Figure 14). 293T cells were transfected with 5µg of each plasmid and 48 hours post transfection mRNA and protein levels were determined as previously described by means of Northern and Western blot analysis respectively.

[0161] Northern blot analysis in cytoplasmic RNA fractions. The blot was probed with a probe complementary to nt 1510-2290 of the codon optimised gene (I) and was re-probed with a probe specific for human ubiquitin (II). Lanes: 1. pSYNGP, 2. pSYN6, 3. pSYN7, 4. pSYN8

Western blot analysis: HIV-1 positive human serum was used to detect the *gag-pol* proteins (I) and an anti-actin antibody was used as an internal control (II). Lanes: Cell lysates: 1. Mock transfected 293T cells, 2. pGP-RRE3 + pCMV-Rev, 3. pSYNGP, 4. pSYN6, 5. pSYN7, 6. pSYN8. Supernatants: 7. Mock transfected 293T cells, 8. pGP-RRE3 + pCMV-Rev, 9. pSYNGP, 10. pSYN6, 11. pSYN7, 12. pSYN8. The protein marker (New England Biolabs) sizes are shown on the side of the gel.

[0162] Figure 14 shows the plasmids used to study the effect of HIV-1 *gag* on the codon optimised gene. The backbone for all constructs was pCI-Neo. Syn *gp*: The codon optimised HIV-1 *gag-pol* gene. HXB2 *gag*: The wild type Hit-1 *gag* gene. HXB2 *gag*: The wild type HIV-1 *gag* gene in the reverse orientation. HXB2 *gag*ΔATG: The wild type HIV-1 *gag* gene without the *gag* ATG. HXB2 *gag*-fr.sh.: The wild type HIV-1 *gag* gene with a frameshift mutation. HXB2 *gag* 625-1503: Nucleotides 625-1503 of the wild type HIV-1 *gag* gene. HXB2 *gag* 1-625: Nucleotides 1-625 of the wild type HIV-1 *gag* gene.

[0163] Figure 15 shows the effect on cytoplasmic RNA of insertion of HIV-1 *gag* upstream of the codon optimised gene. Cytoplasmic RNA was extracted 48 hours post transfection of 293T cells (5µg of each pSYN plasmid was used and 1 µg of pCMV-Rev was co-transfected in some cases). The probe that was used was designed to be complementary to nt 1510-2290 of the codon optimised gene (I). A probe specif for human ubiquitin was used as an internal control (II). Lanes: 1. pSYNGP, 2. pSYN9, 3. pSYN10, 4. pSYN10 + pCMV-Rev, 5. pSYN11, 6. pSYN11 + pCMV-Rev, 7. pCMV-Rev. Lanes: 1. pSYNGP, 2. pSYNGP-RRE, 3. pSYNGP-RRE + pCMV-Rev, 4. pSYN12, 5. pSYN14, 6. pSYN14 + pCMV-Rev, 7. pSYN13, 8. pSYN15, 9. pSYN17, 10. pGP-RRE3, 11. pSYN6, 12. pSYN9, 13. pCMV-Rev.

[0164] Figure 16 shows the effect of LMB on protein production. 293T cells were transfected with 1µg pCMV-Rev and 3µg of pGP-RRE3/pSYNGP/pSYNGP-RRE (+/- 1µg pCMV-Rev). Transfections were done in duplicate. 5 hours post transfection the medium was replaced with fresh medium in the first set and with fresh medium containing 7.5 nM LMB in the second. 20 hours later the cells were lysed and protein production was estimated by Western blot analysis. HIV-1 positive human serum was used to detect the *gag-pol* proteins (A) and an anti-actin antibody was used as an internal control (B). Lanes: 1. pGP-RRE3, 2. pGP-RRE3 + LMB, 3. pGP-RRE3 + pCMV-Rev, 4. pGP-RRE3 + pCMV-Rev + LMB, 5. pSYNGP, 6. pSYNGP + LMB, 7. pSYNGP + pCMV-Rev, 8. pSYNGP + pCMV-Rev + LMB, 9. pSYNGP-RRE, 10. pSYNGP-RRE + LMB, 11. pSYNGP-RRE + pCMV-Rev, 12. pSYNGP-RRE + pCMV-Rev + LMB.

[0165] Figure 17 shows the cytoplasmic RNA levels of the vector genomes. 293T cells were transfected with 10 µg of each vector genome. Cytoplasmic RNA was extracted 48 hours post transfection. 20 µg of RNA were used from each sample for Northern blot analysis. The 700bp probe was designed to hybridise to all vector genome RNAs (see Materials and Methods). Lanes: 1. pH6nZ, 2. pH6nZ + pCMV-Rev, 3. pH6.1nZ, 4. pH6.1nZ + pCMV-Rev, 5. pHS1nZ, 6. pHS2nZ, 7. pHS3nZ, 8. pHS4nZ, 9. pHS5nZ, 10. pHS6nZ, 11. pHS7nZ, 12. pHS8nZ, 13. pCMV-Rev.

[0166] Figure 18 shows transduction efficiency at MOI 1. Viral stocks were generated by co-transfection of each *gag-pol*

expression plasmid (5 or 0.5 µg), 15 µg pH6nZ or pHS3nZ (vector genome plasmid) and 5 µg pHCMVG (VSV envelope expression plasmid) on 293T cells. Virus was concentrated as previously described (45) and transduction efficiency was determined at m.o.i.'s 0.01-1 on HT1080 cells. There was a linear correlation of transduction efficiency and m.o.i. in all cases. An indicative picture at m.o.i. 1 is shown here. Transduction efficiency was >80% with either genome, either *gag-pol* and either high or low amounts of pSYNGP. Titres before concentration (I.U./ml): on 293T cells: A. 6.6×10^5 , B. 7.6×10^5 , C. 9.2×10^5 , D. 1.5×10^5 , on HT1080 cells: A. 6.0×10^4 , B. 9.9×10^4 , C. 8.0×10^4 , D. 2.9×10^4 . Titres after concentration (I.U./ml) on HT1080 cells: A. 6.0×10^5 , B. 2.0×10^6 , C. 1.4×10^6 , D. 2.0×10^5 .

[0167] Figure 21 shows vector titers obtained with differed *gag-pol* constructs. Viral stocks were generated by co-transfection of each *gag-pol* expression plasmid, pH6nZ (vector genome plasmid) and pHCMVG (VSV envelope expression plasmid, 2.5µg for each transfection) on 293T cells. Titres (I.U./ml of virus stock) were measured on 293T cells by counting the number of blue colonies following X-Gal staining 48 hours after transduction. Experiments were performed at least twice and the variation between experiments was less than 15%.

[0168] Figure 22 shows vector titres from the Rev/RRE (-) and (+) genomes. The retroviral vectors were generated as described in the Examples. Titres (I.U./ml of viral stock + SD) were determined in 293T cells.

[0169] Figure 23 shows vector titres from the pHS series of vector genomes. The retroviral vector was generated as described in the Examples. Titres (I.U./ml of viral stock + SD) were determined in 293T cells. Rev is provided from pCMV-Rev. Note that pH6nZ expresses Rev and contains the RRE. None of the other genomes express Rev or contain the RRE. Expression from pSYNGP is Rev independent, whereas it is Rev dependent for pGP-RRE3.

[0170] Figure 24 shows vector titres for the pHS series of vector genomes in the presence or absence of Rev/RRE. The retroviral vector was generated as described in the Examples. 5 µg of vector genome, 5 µg of pSYNGP and 2.5 µg of pHCMVG were used and titres (I.U./ml) were determined in 293T cells. Experiments were performed at least twice and the variation between experiments was less than 15%. Rev is provided from pCMV-Rev (1 µg). Note that pH6nZ expresses Rev and contains the RRE. None of the pHS genomes expresses Rev and only pHS1nZR, pHS3nZR, pHS7nZR and pH6.1nZR contain the RRE. *gag-pol* expression from pSYNGP is Rev independent.

[0171] Figure 26 shows a Western blot of 293T extracts wherein 30:µg of total cellular protein was separated by SDS/Page electrophoresis, transferred to nitro-cellulose and probed with anti EIAV antibodies. The secondary antibody was anti-Horse HRP (Sigma).

[0172] In Figure 38 the titres are shown in lacZ forming units (L.F.U.)/ml. The vectors used are indicated in boxes above the bars.

For ease of reference, we also set out the sequences listed in the accompanying Sequence Listing:

SEQ ID NO:1 shows the sequence of the wild-type *gag-pol* sequence for the strain HXB2 (accession no. K03455);

SEQ ID NO:2 shows the sequence of pSYNGP;

SEQ ID NO:3 shows the sequence of the Envelope gene for HIV-1 MN (Genbank accession no. M17449);

SEQ ID NO:4 shows the sequence of SYNGp-160mn - codon optimised env sequence;

SEQ ID NO:5 shows the sequence of pESYNGP;

SEQ ID NO:6 shows the sequence of LpESYNGP;

SEQ ID NO:7 shows the sequence of pESYNGPRRE;

SEQ ID NO:8 shows the sequence of LpESYNGPRRE;

SEQ ID NO:9 shows the sequence of pONY4.0Z;

SEQ ID NO:10 shows the sequence of pONY8.0Z;

SEQ ID NO:11 shows the sequence of pONY8.1Z;

SEQ ID NO:12 shows the sequence of pONY3.1;

SEQ ID NO: 13 shows the sequence of pCneoERev;

SEQ ID NO:14 shows the sequence of pESYNREV;

SEQ ID NO:15 shows the sequence of codon optimised HIV *gag-pol*;

SEQ ID NO:16 shows the sequence of codon optimised EIAV *gag-pol*;

SEQ ID NO:17 shows the sequence of pIRES1hygESYNGP;

SEQ ID NO:18 shows the sequence of pESDSYNGP; and

SEQ ID NO:19 shows the sequence of pONY8.3G FB29(-).

Example 1 - HIV

Cell lines

[0173] 293T cells (33) and HeLa cells (34) were maintained in Dubecco's modified Eagle's medium containing 10% (v/v) fetal calf serum and supplemented with L-glutamine and antibiotics (penicillin-streptomycin). 293T cells were obtained from D. Baltimore (Rockefeller University).

HIV- 1 proviral clones

[0174] Proviral clones pWI3 (35) and pNL4-3 (36) were used.

Construction of a Packaging System

[0175] In one of the present examples, a modified codon optimised HIV *env* sequence is used (SEQ I.D. No. 4). The corresponding *env* expression plasmid is designated pSYNgp160mn. The modified sequence contains extra motifs not used by (37). The extra sequences were taken from the HIV *env* sequence of strain MN and codon optimised. Any similar modification of the nucleic acid sequence would function similarly as long as it used codons corresponding to abundant tRNAs (38).

Codon optimised HIV-1 *gag-pol* gene

[0176] A codon optimised *gag-pol* gene, shown from nt 1108 to 5414 of SEQ ID NO: 2 was constructed by annealing a series of short overlapping oligonucleotides (approximately 30-40mers with 25% overlap, i.e. approximately 9 nucleotides). Oligonucleotides were purchased from R&D SYSTEMS (R&D Systems Europe Ltd, 4-10 The Quadrant, Barton Lane, Abingdon, OX14 3YS, UK). Codon optimisation was performed using the sequence of HXB-2 strain (AC: K03455) (39). The Kozak consensus sequence for optimal translation initiation (40) was also included. A fragment from base 1222 from the beginning of *gag* until the end of *gag* (1503) was not optimised in order to maintain the frameshift site and the overlap between the *gag* and *pol* reading frames. This was from clone pNL4-3. (When referring to base numbers within the *gag-pol* gene base 1 is the A of the *gag* ATG, which corresponds to base 790 from the beginning of the HXB2 sequence. When referring to sequences outside the *gag-pol* then the numbers refer to bases from the beginning of the HXB2 sequence, where base 1 corresponds to the beginning of the 5' LTR). Some deviations from optimisation were made in order to introduce convenient restriction sites. The final codon usage is shown in Figure 3b, which now resembles that of highly expressed human genes and is quite different from that of the wild type HIV-1 *gag-pol*. The gene was cloned into the mammalian expression vector pCneo (Promega) in the *EcoRI-NotI* sites. The resulting plasmid was named pSYNGP (Figure 20, SEQ ID No 2). Sequencing of the gene in both strands verified the absence of any mistakes. A sequence comparison between the codon optimised and wild type HIV *gag-pol* sequence is shown in Figure 8.

Rev/RRE constructs

[0177] The HIV-1 RRE sequence (bases 7769-8021 of the HXB2 sequence) was amplified by PCR from pWI3 proviral clone with primers bearing the *NotI* restriction site and was subsequently cloned into the *NotI* site of pSYNGP. The resulting plasmids were named pSYNGP-RRE (RRE in the correct orientation) and pSYNGP-ERR (RRE in the reverse orientation).

Pseudotyped viral particles

[0178] In one form of the packaging system a synthetic *gag-pol* cassette is coexpressed with a heterologous envelope coding sequence. This could be for example VSV-G (44, 45), amphotropic MLV env (46, 47) or any other protein that would be incorporated into the HIV or EIAV particle (48). This includes molecules capable of targeting the vector to specific tissues.

HIV-1 Vector genome constructs

[0179] pH6nZ is derived from pH4Z (49) by the addition of a single nucleotide to place an extra guanine residue that was missing from pH4Z at the 5' end of the vector genome transcript to optimise reverse transcription. In addition the gene coding for β -galactosidase (LacZ) was replaced by a gene encoding for a nuclear localising β -galactosidase. (We are grateful to Enca Martin-Rendon and Said Ismail for providing pH6nZ). In order to construct Rev(-) genome constructs the following modifications were made : a) A 1.8 kb *PstI* - *PstI* fragment was removed from pH6nZ, resulting in plasmid pH6.1nZ and b) an *EcoNI* (filled) - *SphI* fragment was substituted with a *SpeI* (filled) - *SphI* fragment from the same plasmid (pH6nZ), resulting in plasmid pH6.2nZ. In both cases sequences within *gag* (nt 1-625) were retained, as they have been shown to play a role in packaging (93). *Rev*, RRE and any other residual *env* sequences were removed. pH6.2nZ further contains the *env* splice acceptor, whereas pH6.1nZ does not.

[0180] A series of vectors encompassing further *gag* deletions plus or minus a mutant major splice donor (SD) (GT to CA mutation) were also derived from pH6Z. These were made by PCR with primers bearing a *NarI* (5' primers) and an *SpeI* (3' primers) site. The PCR products were inserted into pH6Z at the *NarI* - *SpeI* sites. The resulting vectors were named pHS1nZ (containing HIV-1 sequences up to *gag* 40), pHS2nZ (containing HIV-1 sequences up to *gag* 260), pHS3nZ (containing HIV-1 sequences up to *gag* 360), pHS4nZ (containing HIV-1 sequences up to *gag* 625), pHS5nZ (same as pHS1nZ but with a mutant SD), pHS6nZ (same as pHS2nZ but with a mutant SD), pHS7nZ (same as pHS3nZ but with a mutant SD) and pHS8nZ (same as pHS4nZ but with a mutant SD).

[0181] In addition, the RRE sequence (nt 7769-8021 of the HXB2 sequence) was inserted in the *SpeI* (filled) site of pH6.1nZ, pHS1nZ, pHS3nZ and pHS7nZ resulting in plasmids pH6.1nZR, pHS1nZR, pHS3nZR and pHS7nZR respectively.

[0182] Other modifications to the genome have been made including the generation of a SIN vector (by deletion of part of the 3' U3), the replacement of the LTRs with those from MLV or replacement of part of the 3' U3 with the MLV U3 region.

Transient transfections, transductions and determination of viral titres

[0183] These were performed as previously described (49, 50). Briefly, 293T cells were seeded on 6cm dishes and 24 hours later they were transiently transfected by overnight calcium phosphate treatment. The medium was replaced 12 hours post-transfection and unless otherwise stated supernatants were harvested 48 hours post-transfection, filtered (through 0.22 or 0.45 μ m filters) and titered by transduction of 293T cells. For this reason supernatant at appropriate dilutions of the original stock was added to 293T cells (plated onto 6 or 12 well plates 24 hours prior to transduction). 8 μ g/ml Polybrene (Sigma) was added to each well and 48 hours post transduction viral titres were determined by X-gal staining.

Luminescent β -galactosidase (β -gal) assays

[0184] These were performed on total cell extracts using a luminescent β -gal reporter system (CLONTECH). Untransfected 293T cells were used as negative control and 293T cells transfected with pCMV- β -gal (CLONTECH) were used as positive control.

RNA analysis

[0185] Total or cytoplasmic RNA was extracted from 293T cells by using the RNeasy mini kit (QUIAGEN) 36-48 hours post-transfection. 5-10 µg of RNA was subjected to Northern blot analysis as previously described (51). Correct fractionation was verified by staining of the agarose gel. A probe complementary to bases 1222-1503 of the *gag-pol* gene was amplified by PCR from HIV-1 pNL4-3 proviral clone and was used to detect both the codon optimised and wild type *gag-pol* mRNAs. A second probe, complementary to nt 1510-2290 of the codon optimised gene was also amplified by PCR from plasmid pSYNGP and was used to detect the codon optimised genes only. A 732 bp fragment complementary to all vector genomes used in this study was prepared by an *SpeI*-*AvrII* digestion of pH6nZ. A probe specific for ubiquitin (CLONTECH) was used to normalise the results. All probes were labelled by random labelling (STRATAGENE) with α -³²P dCTP (Amersham). The results were quantitated by using a Storm PhosphorImager (Molecular Dynamics) and shown in Figure 12. In the total cellular fractions the 47S rRNA precursor could be clearly seen, whereas it was absent from the cytoplasmic fractions. As expected (52), Rev stimulates the cytoplasmic accumulation of wild type *gag-pol* mRNA (lanes 1c and 2c). RNA levels were 10-20 fold higher for the codon optimised gene compared to the wild type one, both in total and cytoplasmic fractions (compare lanes 3t-2t, 3c-2c, 10c-8c). The RRE sequence did not significantly destabilise the codon optimised RNAs since RNA levels were similar for codon optimised RNAs whether or not they contained the RRE sequence (compare lanes 3 and 5). Rev did not markedly enhance cytoplasmic accumulation of the codon optimised *gag-pol* mRNAs, even when they contained the RRE sequence (differences in RNA levels were less than 2-fold, compare lanes 3-4 or 5-6).

[0186] It appeared from a comparison of Figures 10 and 12 that all of the increase in protein expression from *syngp* could be accounted for by the increase in RNA levels. In order to investigate whether this was due to saturating levels of RNA in the cell, we transfected 0.1, 1 and 10 µg of the wild type or codon optimised expression vectors into 293T cells and compared protein production. In all cases protein production was 10-fold higher for the codon optimised gene for the same amount of transfected DNA, while increase in protein levels was proportional to the amount of transfected DNA for each individual gene. It seems likely therefore that the enhanced expression of the codon optimised gene can be mainly attributed to the enhanced RNA levels present in the cytoplasm and not to increased translation.

Protein analysis

[0187] Total cell lysates were prepared from 293T cells 48 hours post-transfection (unless otherwise stated) with an alkaline lysis buffer. For extraction of proteins from cell supernatants the supernatant was first passed through a 0.22µm filter and the vector particles were collected by centrifugation of 1 ml of supernatant at 21,000 g for 30 minutes. Pellets were washed with PBS and then re-suspended in a small volume (2-10 µl) of lysis buffer. Equal protein amounts were separated on a SDS 10-12% (v/v) polyacrylamide gel. Proteins were transferred to nitrocellulose membranes which were probed sequentially with a 1:500 dilution of HIV-1 positive human serum (AIDS Reagent Project, ADP508, Panel E) and a 1:1000 dilution of horseradish peroxidase labelled anti-human IgG (Sigma, A0176). Proteins were visualised using the ECL or ECL-plus western blotting detection reagent (Amersham). To verify equal protein loading, membranes were stripped and re-probed with a 1:1000 dilution of anti-actin antibody (Sigma, A2066), followed by a 1:2000 dilution of horseradish peroxidase labelled anti-rabbit IgG (Vector Laboratories, PI-1000).

Expression of *gag-pol* gene products and vector particle production

[0188] The wild type *gag-pol* (pGP-RRE3 B Figure 19) (49), and codon optimised expression vectors (pSYNGP, pSYNGP-RRE and pSYNGP-ERR) were transiently transfected into 293T cells. Transfections were performed in the presence or absence of a Rev expression vector, pCMV-Rev (53), in order to assess Rev-dependence for expression. Western blot analysis was performed on cell lysates and supernatants to assess protein production. The results are shown in Figure 10. As expected (54), expression of the wild type gene is observed only when Rev is provided *in trans* (lanes 2 and 3). In contrast, when the codon optimised *gag-pol* was used, there was high level expression in both the presence and absence of Rev (lanes 4 and 5), indicating that in this system there was no requirement for Rev. Protein levels were higher for the codon optimised gene than for the wild type *gag-pol* (compare lanes 4-9 with lane 3). The difference was more evident in the cell supernatants (approximately 10-fold higher protein levels for the codon optimised gene compared to the wild type one, quantitated by using a PhosphorImager) than in the cell lysates.

[0189] In previous studies where the RRE has been included in *gag-pol* expression vectors that had been engineered to remove INS sequences, inclusion of the RRE lead to a decrease in protein levels, that was restored by providing Rev *in trans* (55). In our hands, the presence of the RRE in the fully codon optimised *gag-pol* mRNA did not affect protein levels and provision of Rev *in*

trans did not further enhance expression (lanes 6 and 7).

[0190] In order to compare translation rates between the wild type and codon optimised gene, protein production from the wild type and codon optimised expression vector was determined at several time intervals post transfection into 293T cells. Protein production and particle formation was determined by Western blot analysis and the results are shown in Figure 11. Protein production and particle formation was 10-fold higher for the codon optimised *gag-pol* at all time points.

[0191] To further determine whether this enhanced expression that was observed with the codon optimised gene was due to better translation or due to effects on the RNA, RNA analysis was carried out.

The efficiency of vector production using the codon optimised *gag-pol* gene

[0192] To determine the effects of the codon optimised *gag-pol* on vector production, we used an HIV vector genome, pH6nZ and the VSV-G envelope expression plasmid pHCMVG (113), in combination with either pSYNGP, pSYNGP-RRE, pSYNGP-ERR or pGP-RRE3 as a source for the *gag-pol* in a plasmid ratio of 2 : 1 : 2 in a 3 plasmid co-transfection of 293T cells (49). Whole cell extracts and culture supernatants were evaluated by Western blot analysis for the presence of the *gag* and *gag-pol* gene products. Particle production was, as expected (Figure 10), 5-10 fold higher for the codon optimised genes when compared to the wild type.

[0193] To determine the effects of the codon optimised *gag-pol* gene on vector titres, several ratios of the vector components were used. The results are shown in Figure 21. Where the *gag-pol* was the limiting component in the system (as determined by the drop in titres observed with the wild type gene), titres were 10-fold higher for the codon optimised vectors. This is in agreement with the higher protein production observed for these vectors, but suggests that under normal conditions of vector production *gag-pol* is saturating and the codon optimisation gives no maximum yield advantage.

The effect of HIV-1 *gag* INS sequences on the codon optimised gene is position dependent

[0194] It has previously been demonstrated that insertion of wild type HIV-1 *gag* sequences downstream of other RNAs, e.g. HIV-1 *tat* (56), HIV-1 *gag* (55) or CAT (57) can lead to a dramatic decrease in steady state mRNA levels, presumably as a result of the INS sequences. In other cases, e.g. for β -globin (58), it was shown that the effect was splice site dependent. Cellular AREs (AU-rich elements) that are found in the 3' UTR of labile mRNAs may confer mRNA destabilisation by inducing cytoplasmic deadenylation of the transcripts (59). To test whether HIV-1 *gag* INS sequences would destabilise the codon optimised RNA, the wild-type HIV-1 *gag* sequence, or parts of it (nt 1-625 or nt 625-1503), were amplified by PCR from the proviral clone pW13. All fragments were blunt ended and were inserted into pSYNGP or pSYNGP-RRE at either a blunted *EcoR1* or *NotI* site (upstream or downstream of the codon optimised *gag-pol* gene respectively). As controls the wt HIV-1 *gag* in the reverse orientation (as INS sequences have been shown to act in an orientation dependent manner, (57) (pSYN7) and *lacZ*, excised from plasmid pCMV- β gal (CLONTECH) (in the correct orientation) (pSYN8) were also inserted in the same site. Contrary to our expectation, as shown in Figure 13, the wild type HIV-1 *gag* sequence did not appear to significantly affect RNA or protein levels of the codon optimised gene. We further constructed another series of plasmids (by PCR and from the same plasmids) where the wild type HIV-1 *gag* in the sense or reverse orientation, subfragments of *gag* (nt 1-625 or nt 625-1503), the wild type HIV-1 *gag* without the ATG or with a frameshift mutation 25 bases downstream of the ATG, or nt 72-1093 of *LacZ* (excised from plasmid pH6Z), or the first 1093 bases of *lacZ* with or without the ATG were inserted upstream of the codon optimised HIV-1 *gag-pol* gene in pSYNGP and/or pSYNGP-RRE (pSYN9-pSYN22, Figure 14). Northern blot analysis showed that insertion of the wild type HIV-1 *gag* gene upstream of the codon optimised HIV-1 *gag-pol* (pSYN9, pSYN10) lead to diminished RNA levels in the presence or absence of Rev/RRE (Figure 15A, lanes 1-4 and Figure 15B, lanes 1+12). The effect was not dependent on translation as insertion of a wild type HIV-1 *gag* lacking the ATG or with a frameshift mutation (pSYN12, pSYN13 and pSYN14) also diminished RNA levels (Figure 15B, lanes 1-7). Western blot analysis verified that there was no HIV-1 *gag* translation product for pSYN12-14. However, it is possible that, as the wt HIV-1 *gag* exhibits such an adverse codon usage, it may act as a non-translatable long 5' leader for *syngp*, and if this is the case, then the ATG mutation should not have any effects.

[0195] Insertion of smaller parts of the wild type HIV-1 *gag* gene (pSYN15 and pSYN17) also lead to a decrease in RNA levels (Figure 15B, lanes 1-3 and 8-9), but not to levels as low as when the whole *gag* sequence was used (lanes 1-3, 4-7 and 8-9 in Figure 15B). This indicates that the effect of INS sequences is dependent on their size. Insertion of the wild type HIV-1 *gag* in the reverse orientation (pSYN11) had no effect on RNA levels (Figure 15A, lanes 1 and 5-6). However a splicing event seemed to take place in that case, as indicated by the size of the RNA (equal to the size of the codon optimised *gag-pol* RNA) and by the

translation product (gag-pol, in equal amounts compared to pSYNGP, as verified by Western blot analysis).

[0196] These data indicate therefore that wild type HIV-1 *gag* instability sequences act in a position and size dependent manner, probably irrespective of translation. It should also be noted that the RRE was unable to rescue the destabilised RNAs through interaction with Rev.

Construction of an HIV-1 based vector system that lacks all the accessory proteins

[0197] Until now several HIV-1 based vector systems have been reported that lack all accessory proteins but Rev (49, 60). We wished to investigate whether the codon optimised gene would permit the construction of an HIV-1 based vector system that lacks all accessory proteins. We initially deleted *rev*/RRE and any residual *env* sequences, but kept the first 625 nucleotides of *gag*, as they have been shown to play a role in efficient packaging (61). Two vector genome constructs were made, pH6.1nZ (retaining only HIV sequences up to nt 625 of *gag*) and pH6.2nZ (same as pH6.1nZ, but also retaining the *env* splice acceptor). These were derived from a conventional HIV vector genome that contains RRE and expresses Rev (pH6nZ). Our 3-plasmid vector system now expressed only HIV-1 gag-pol and the VSV-G envelope proteins. Vector particle titres were determined as described in the previous section. A ratio of 2 : 2 : 1 of vector genome (pH6Z or pH6.1nZ or pH6.2nZ) : gag-pol expression vector (pGP-RRE3 or pSYNGP) : pHCMV-G was used. Transfections were performed in the presence or absence of pCMV-Rev, as gag-pol expression was still Rev dependent for the wild type gene. The results are summarised in Figure 22 and indicate that an HIV vector could be produced in the total absence of Rev, but that maximum titres were compromised at 20-fold lower than could be achieved in the presence of Rev. As gag-pol expression should be the same for pSYNGP with pH6nZ or pH6.1nZ or pH6.2nZ (since it is Rev independent), as well as for pGP-RRE3 when Rev is provided *in trans*, we suspected that the vector genome retained a requirement for Rev and was therefore limiting the titres. To confirm this, Northern blot analysis was performed on cytoplasmic RNA prepared from cells transfected with pH6nZ or pH6.1nZ in the presence or absence of pCMV-Rev. As can be seen in Figure 17, lanes 1-4, the levels of cytoplasmic RNA derived from pH6nZ were 5-10 fold higher than those obtained with pH6.1nZ (compare lanes 1-2 to lanes 3-4). These data support the notion that RNA produced from the vector genome requires the Rev/RRE system to ensure high cytoplasmic levels. This may be due to inefficient nuclear export of the RNA, as INS sequences residing within *gag* were still present.

[0198] Further deletions in the *gag* sequences of the vector genome might therefore be necessary to restore titres. To date efficient packaging has been reported to require 360 (62) or 255 (63) nucleotides of *gag* in vectors that still retain *env* sequences, or about 40 nucleotides of *gag* in a particular combination of splice donor mutation, *gag* and *env* deletions (64, 63). In an attempt to remove the requirement for Rev/RRE in our vector genome without compromising efficient packaging we constructed a series of vectors derived from pH6nZ containing progressively larger deletions of HIV-1 sequences (only sequences upstream and within *gag* were retained) plus and minus a mutant major splice donor (SD) (GT to CA mutation). Vector particle titres were determined as before and the results are summarized in Figure 23. As can be seen, deletion of up to nt 360 in *gag* (vector pHS3nZ) resulted in an increase in titres (compared to pH6.1nZ or pH6.2nZ) and only a 5-fold decrease (titres were $1.3-1.7 \times 10^5$) compared to pH6nZ. Further deletions resulted in titres lower than pHS3nZ and similar to pH6.1nZ. In addition, the SD mutation did not have a positive effect on vector titres and in the case of pHS3nZ it resulted in a 10-fold decrease in titres (compare titres for pHS3nZ and pHS7nZ in Figure 23). Northern blot analysis on cytoplasmic RNA (Figure 17, lanes 1 and 5-12) showed that RNA levels were indeed higher for pH6nZ, which could account for the maximum titres observed with this vector. RNA levels were equal for pHS1nZ (lane 5), pHS2nZ (lane 6) and pHS3nZ (lane 7) whereas titres were 5-8 fold higher for pHS3nZ. It is possible that further deletions (than that found in pHS3nZ) in *gag* might result in less efficient packaging (as for HIV-1 the packaging signal extends in *gag*) and therefore even though all 3 vectors produce similar amounts of RNA only pHS3nZ retains maximum packaging efficiency. It is also interesting to note that the SD mutation resulted in increased RNA levels in the cytoplasm (compare lanes 6 and 10, 7 and 11 or 8 and 12 in Figure 17) but equal or decreased titres (Figure 23). The GT dinucleotide that was mutated is in the stem of SL2 of the packaging signal (65). It has been reported that SL2 might not be very important for HIV-1 RNA encapsidation (65, 66), whereas SL3 is of great importance (67). Folding of the wild type and SD-mutant vector sequences with the RNAdraw software program revealed that the mutation alters significantly the secondary structure of the RNA and not only of SL2. It is likely therefore that although the SD mutation enhances cytoplasmic RNA levels it does not increase titres as it alters the secondary structure of the packaging signal.

[0199] To investigate whether the titre differences that were observed with the Rev minus vectors were indeed due to Rev dependence of the genomes, the RRE sequence (nt 7769 - 8021 of the HXB2 sequence) was inserted in the *SpeI* site (downstream of the *gag* sequence and just upstream of the internal CMV promoter) of pH6.1nZ, pHS1nZ, pHS3nZ and pHS7nZ, resulting in plasmids pH6.1nZR, pHS1nZR, pHS3nZR and pHS7nZR respectively. Vector particle titres were determined with pSYNGP and pHCMVG in the presence or absence of Rev (pCMV-Rev) as before and the results are summarised in Figure 24. In

the absence of Rev titres were further compromised for pH6.1nZR (7-fold compared to pH6.1nZ), pHS3nZR (6-fold compared to pHS3nZ) and pHS7nZR (2.5-fold compared to pHS7nZ). This was expected, as the RRE also acts as an instability sequence (68) and so it would be expected to confer Rev-dependence. In the presence of Rev titres were restored to the maximum titres observed for pH6nZ in the case of pHS3nZR (5×10^5) and pH6.1nZR (2×10^5). Titres were not restored for pHS7nZR in the presence of Rev. This supports the hypothesis that the SD mutation in pHS7nZ affects the structure of the packaging signal and thus the packaging ability of this vector genome, as in this case Rev may be able to stimulate vector genome RNA levels, as for pHS3nZR and pH6.1nZR, but it can not affect the secondary structure of the packaging signal. For vector pHS1nZ inclusion of the RRE did not lead to a decrease in titres. This could be due to the fact that pHS1nZ contains only 40 nucleotides of *gag* sequences and therefore even with the RRE the size of instability sequences is not higher than for pHS2nZ that gives equal titres to pHS1nZ. Rev was able to partially restore titres for pHS1nZR (10-fold increase when compared to pHS1nZ and 8-fold lower than pH6nZ) but not fully as in the case of pHS3nZ. This is also in agreement with the hypothesis that 40 nucleotides of HIV-1 *gag* sequences might not be sufficient for efficient vector RNA packaging and this could account for the partial and not complete restoration in titres observed with pHS1nZR in the presence of Rev.

[0200] In addition, end-point titres were determined for pHS3nZ and pH6nZ with pSYNGP in HeLa and HT1080 human cell lines. In both cases titres followed the pattern observed in 293T cells, with titres being 2-3 fold lower for pHS3nZ than for pH6nZ (See Figure 10). Finally, transduction efficiency of vector produced with pHS3nZ or pH6nZ and different amounts of pSYNGP or pGP-RRE3 at different m.o.i.s (and as high as 1) was determined in HT1080 cells. This experiment was performed as the high level *gag-pol* expression from pSYNGP may result in interference by genome-empty particles at high vector concentrations. As expected for VSVG pseudotyped retroviral particles (69) transduction efficiencies correlated with the m.o.i.s, whether high or low amounts of pSYNGP were used and with pH6nZ or pHS3nZ. For m.o.i. 1 transduction efficiency was approximately 50-60% in all cases (Figure 18). The above data indicate that no interference due to genome-empty particles is observed in this experimental system.

The codon optimised *gag-pol* gene does not use the exportin-1 nuclear export pathway

[0201] Rev mediates the export of unspliced and singly spliced HIV-1 mRNAs via the nuclear export receptor exportin-1 (CRM1) (70, 71, 72, 73, 74). Leptomycin B (LMB) has been shown to inhibit leucine-rich NES mediated nuclear export by disrupting the formation of the exportin-1/NES/RanGTP complex (75, 72). In particular, LMB inhibits nucleo-cytoplasmic translocation of Rev and Rev-dependent HIV mRNAs (76). To investigate whether exportin-1 mediates the export of the codon optimised *gag-pol* constructs, the effect of LMB on protein production was tested. Western blot analysis was performed on cell lysates from cells transfected with the *gag-pol* constructs (+/- pCMV-Rev) and treated or not with LMB (7.5 nM, for 20 hours, beginning treatment 5 hours post-transfection). To confirm that LMB had no global effects on transport, the expression of β -gal from the control plasmid pCMV- β Gal was also measured. An actin internal control was used to account for protein variations between samples. The results are shown in Figure 16. As expected (76), the wild type *gag-pol* was not expressed in the presence of LMB (compare lanes 3 and 4), whereas LMB had no effect on protein production from the codon optimised *gag-pol*, irrespective of the presence of the RRE in the transcript and the provision of Rev *in trans* (compare lanes 5 and 6, 7 and 8, 9 and 10, 11 and 12, 5-6 and 11-12). The resistance of the expression of the codon-optimised *gag-pol* to inhibition by LMB indicates that the exportin-1 pathway is not used and therefore an alternative export pathway must be used. This offers a possible explanation for the Rev independent expression. The fact that the presence of a nonfunctional Rev/RRE interaction did not affect expression implies that the RRE does not necessarily act as an inhibitory (e.g. nuclear retention) signal *per se*, which is in agreement with previous observations (5, 58).

[0202] In conclusion, this is the first report of an HIV-1 based vector system, composed of pSYNGP, pHS3nZ and pHCMVG, where significant vector production can be achieved in the absence of all accessory proteins. These data indicate that in order to achieve maximum titres the HIV vector genome must be configured to retain efficient packaging and that this requires the retention of *gag* sequences and a splice donor. By reducing the *gag* sequence to 360 nt in pHS3nZ and combining this with pSYNGP it is possible to achieve titre of at least 10^5 I.U./ml that is only 5-fold lower than the maximum levels achieved in the presence of Rev.

Example 2 - EIAV (not according to the invention)

Codon-optimised EIAV *gag-pol* expression cassettes

[0203] We also examined if the codon-optimisation process would alter the properties of the *gag-pol* gene of the non-primate lentivirus EIAV. The sequence of the codon-optimised gene is shown from nt1103 to 5760 of SEQ ID NO:5 (Figure 9). The wild type and the codon-optimised sequences are denoted WT and CO, respectively. The codon usage was changed to that of highly expressed mammalian genes. pESYNGP (Figure 27 and SEQ ID NO:5) was made by transferring an *XbaI*-*NotI* fragment from a plasmid containing a codon-optimised EIAV *gag/pol* gene, synthesised by Operon Technologies Inc., Alameda, CA, into pCneo (Promega). The gene was supplied in a proprietary plasmid backbone, GeneOp. The fragment transferred to pCneo includes sequences flanking the codon-optimised EIAV *gag/pol* ORF: **tctagaGAATTCGCCACCATG- EIAV gag/pol-TGAACCCGGGgcggcgcg**. The ATG start and TGA stop codons are shown in bold and the recognition sequences for *XbaI* and *NotI* sites in lower case.

[0204] The expression of Gag/Pol from the codon-optimised gene was assessed with respect to that from various wild type EIAV *gag/pol* expression constructs by transient transfection of HEK 293T cells (Figure 25). Transfections were carried out using the calcium phosphate technique, using equal moles of each Gag/Pol expression plasmid together with a plasmid which expressed EIAV Rev either from the wild type sequence or from a codon-optimised version of the gene: pCneoEREV (WO 99/32646) (Figure 35 and SEQ ID NO:13) or pESYNREV (Figure 36 and SEQ ID NO:14), respectively. pESYNREV is a pCneo-based plasmid (Promega) which was made by introducing the *EcoRI* to *SalI* fragment from a synthetic EIAV REV plasmid, made by Operon Technologies Alameda, CA. The plasmid backbone was the proprietary plasmid GeneOp in which was inserted a codon-optimised EIAV REV gene flanked by *EcoRI* and *SalI* recognition sequences and a Kozak consensus sequence to drive efficient translation of the gene. The mass of DNA on each transfection was equalised by addition of pCneo plasmid. In transfections in which a Rev expression plasmid was omitted, a similar mass of pCneo (Promega) was used instead (lanes labelled pCneo). Cytoplasmic extracts were prepared 48 hours post transfection and 15µg amounts of protein were fractionated by SDS-PAGE and then transferred to Hybond ECL. The Western blot was probed with a polyclonal antisera from an EIAV-infected horse and then with a secondary antibody, anti-horse horse-radish peroxidase conjugate. Development of the blot was carried out using the ECL kit (Amersham). Positive controls for the blotting and development procedure, and cytoplasmic extract from untransfected HEK 293T cells are as indicated. The positions of various EIAV proteins are indicated.

[0205] Expression from wild type *gag/pol* was achieved from various plasmids (see Figure 25). pONY3.2T is a derivative of pONY3.1 (WO99/32646)(Figure 34 and SEQ ID NO:12) in which mutations which ablate expression of Tat and S2 have been made. In addition, the EIAV sequence is truncated downstream of the second exon of *rev*. Specifically, expression of Tat is ablated by an 83nt deletion in exon 2 of *tat* which corresponds with respect to the wild type EIAV sequence, Acc. No. U01866, to deletion of nt 5234-5316 inclusive. S2 ORF expression is ablated by a 51 nt deletion, corresponding to nt 5346-5396 of Acc. No. U01866. The EIAV sequence is deleted downstream of a position corresponding to nt 7815 of Acc. No. U01866. These alterations do not alter *rev*, hence expression of this gene is expressed as for pONY3.1. pONY3.2 OPTI is a derivative of pONY3.1 which has the same deletions for ablation of Tat and S2 expression as described above. In addition, the first 372nt of *gag* have been 'codon-optimised' for expression in human cells. The sequence of the wild type and codon-optimised sequences present in pONY3.2OPTI in this region are compared in Figure 43. Base differences between the sequences are indicated. The region which was codon-optimised represents the region of overlap between the vector and wild-type *gag/pol* expression constructs. Reduction of homology within this region would be expected to improve the safety profile of the vector system due to the reduced chances of recombination between the vector genome and the *gag/pol* transcripts. 3.2 OPTI-lhyg is a derivative of 3.2 OPTI in which the *SnaBI*-*NotI* fragment of 3.2 OPTI is transferred to pIRES1hygro (Clontech) prepared for ligation by digestion with the same sites. The *gag/pol* gene is thus placed upstream of the IRES hygromycin phosphotransferase. Of note is the fact that the resulting construct contains the intron from pCneo, not from pIRES1hygro. pEV53B is a derivative of PEV53A (WO 98/51810) in which the EIAV-derived sequence upstream of the Gag initiation codon is reduced to include only the major splice donor and surrounding sequences: CAG/GTAAGATG, where the Gag initiation codon is shown in bold face.

[0206] The results (Figure 26) shown the Rev-dependence of Gag/Pol expression from pHORSE3.1 (WO 99/32646), which has an EIAV derived leader sequence starting just downstream of the primer binding site and an RRE placed downstream of *gag/pol* composed of the two EIAV sequences reported to have RRE activity. Expression was enhanced by the same amount when Rev expression was driven by wild type (pCneoEREV) (Figure 35) or codon-optimised (pESYNREV) (Figure 36) genes. This result confirms the functionality of the codon-optimised Rev expression plasmid.

[0207] In contrast to expression of Gag/Pol from pONY3.1, expression from pESYNGP was not influenced by the presence of Rev, however it was slightly lower than from pONY3.1 or pON3.2T. Expression from pESYNGPRRE (Figure 30 and SEQ ID NO:7), in which the EIAV RRE sequence present in pHORSE3.1 is placed downstream of *gag/pol*, appeared slightly lower than from pESYNGP. The levels of expression from 3.2 OPTI and 3.2OPTI-lhyg were significantly lower than from pESYNGP or pONY3.1, even in the presence of Rev. This result suggested that there may be determinants of Gag/Pol expression within the first 372nt of the *gag* and showed that 3.2 OPTI was unlikely to be useful as a basis for EIAV vector production. Furthermore it demonstrates that codon-optimisation of only certain regions of the whole *gag/pol* gene may not lead to high levels of Rev-independent

expression.

[0208] We have previously demonstrated (43) that the 5' leader (121bp upstream of the ATG start codon) and the RRE sequence (43) are important for high expression of the wild type EIAV *gag-pol*. Three constructs were made that contained either the leader sequence (LpESYNGP), the leader and RRE sequences (LpESYNGPRRE) or the RRE sequence (pESYNGPRRE). The sequences of these constructs are shown in SEQ ID NOS:6-8 and Figures 28-30. They were transfected into 293T cells in either the presence or absence of Rev expression plasmid. The cell supernatant was then measured for reverse transcriptase activity (RT), using a conventional RT assay, to evaluate which construct generated the highest amount of *gag-pol* mRNA. The results are shown in Figures 39 and 40. It is clear from these results that the 5' leader leads to an increase in RT activity. The ability of these Gag/Pol expression constructs to support formation of infectious vector particles was also tested by transient transfection of HEK 293 cells. The results of this analysis show that all of the constructs could provide functional EIAV Gag/Pol, and show the Rev dependence of titre with the pONY8.0Z vector genome plasmid, which does not encode any EIAV proteins (Figure 41).

[0209] The ability of pESYNGP to act in concert with a minimal EIAV vector genome plasmid pONY8.1Z (Figure 33, SEQ ID NO:11) was evaluated (Figure 42). The result shows that the titres obtained with pESYNGP and pONY8.1Z are about 10-fold lower than from pONY3.1 and pONY8.1Z. This reduced titre reflects the lack of Rev protein in the system rather than a deficiency of Gag/Pol production which we have already shown is independent of Rev expression.

[0210] Expression of EIAV Gag/Pol was also tested from pESDSYNGP (Figure 50 and SEQ ID NO:18) in which the Kozak consensus sequence of Gag is replaced by the natural EIAV splice donor. pESDSYNGP was made from pESYNGP by exchange of the 306bp *EcoRI-NheI* fragment, which runs from just upstream of the start codon for *gag/pol* to approximately 300 base pairs inside the *gag/pol* ORF with a 308bp *EcoRI-NheI* fragment derived by digestion of a PCR product made using pESYNGP as template and using the following primers: SD FOR [GGCTAGAGAATTCCAGGTAAGATGGGCGATCCCCTCACCTGG] and SD REV [TTGGGTACTCCTCGCTAGGTTC]. This manipulation replaces the Kozak consensus sequence upstream of the ATG in pESYNGP with the splice donor found in EIAV. The sequence between the *EcoRI* site and the ATG of *gag/pol* is thus CAGGTAAG, exactly as found in the natural viral sequence. Therefore the mRNA is deleted with respect to sequences upstream but not downstream of the splice donor. The performance of pESDSYNGP was assessed relative to pESYNGP and other expression plasmids by measurement of reverse transcriptase activity in supernatants from transiently transfected HEK 293T cells using a Taqman-based version of the product enhanced reverse transcriptase (PERT) assay. In this method, reverse transcriptase associated with vector particles is released by mild detergent treatment and used to synthesize cDNA using MS2 bacteriophage RNA as template. MS2 RNA template and primer are present in excess hence the amount of cDNA is proportional to the amount of RT released from the particles. Therefore, the amount of cDNA synthesised is proportional to the number of particles. MS2 cDNA is then quantitated using Taqman technology. The assay is carried out on test samples in parallel with a vector stock of known titre and estimated particle content. The use of the standard allows creation of a 'standard curve' and allows the relative RT content of various samples to be calculated. The results of this analysis are shown in Figure 49. The results show that Gag/Pol expression is virtually identical from pESYNGP and pESDSYNGP. The results also indicate that expression is not significantly enhanced by Rev. The activity of the Rev expression plasmid is confirmed by the result obtained with pHORSE +, in which there is an RRE downstream of the wild type EIAV *gag/pol*, and that shows a 6-fold enhancement of expression in the presence of Rev. We also noted that the expression from pHORSE was enhanced 3-fold in the presence of Rev. Since this construct has no RRE it suggests that Rev may be having a non-specific enhancing effect on expression, possibly as a result of being expressed at high levels in this experimental system.

[0211] The ability of pESYNGP to participate in the formation of infectious viral vector particles, when co-transfected with plasmids for the vector genome and envelope was assessed by transient transfection of HEK 293T, as described previously (49, 50). Briefly, 293T cells were seeded on 6cm dishes (1.2×10^6 /dish) and 24 hours later they were transfected by the calcium phosphate procedure. The medium was replaced 12 hours post-transfection and supernatants were harvested 48 hours post-transfection, filtered (0.45 μ m filters) and titered by transduction of D17, canine osteosarcoma cells, in the presence of 8 μ g/ml Polybrene (Sigma). Cells were seeded at 0.9×10^5 /well in 12 well plates 24 hours prior to use in titration assays. Dilutions of supernatant were made in complete media (DMEM/10%FBS) and 0.5ml aliquots plated out onto the D17 cells. 4 hours after addition of the vector the media was supplemented, with a further 1 ml of media. Transduction was assessed by X-gal staining of cells 48 hours after addition of viral dilutions.

[0212] The vector genomes used for these experiments were pONY4.0Z (Figure 31 and SEQ ID NO:9) and pONY8.0Z (Figure 32 and SEQ ID NO:10).

[0213] pONY4.0Z (WO 99/32646) was derived from pONY2.11Z by replacement of the U3 region in the 5'LTR with the

cytomegalovirus immediate early promoter (pCMV). This was carried out in such a way that the first base of the transcript derived from this CMV promoter corresponds to the first base of the R region. This manipulation results in the production of high levels of vector genome in transduced cells, particularly HEK 293T cells, and has been described previously (50). pONY4.0Z expresses all EIAV proteins except for envelope, expression of which is ablated by a deletion of 736nt between the *HindIII* sites present in *env*.

[0214] pONY8.0Z was derived from pONY4.0Z by introducing mutations which 1) prevented expression of TAT by an 83nt deletion in the exon 2 of *tat*) prevented S2 ORF expression by a 51 nt deletion 3) prevented REV expression by deletion of a single base within exon 1 of *rev* and 4) prevented expression of the N-terminal portion of *gag* by insertion of T in ATG start codons, thereby changing the sequence to ATTG from ATG. With respect to the wild type EIAV sequence Acc. No. U01866 these correspond to deletion of nt 5234-5316 inclusive, nt 5346-5396 inclusive and nt 5538. The insertion of T residues was after nt 526 and 543.

[0215] The results of this analysis are shown tabulated in Figure 37, and graphically in Figure 38. Transfections were carried out with only 3 plasmids (vector genome, *gag/pol* expression plasmid and VSV-G expression plasmid) - diagonal lined bars, or with four plasmids, which included the previous set of plasmids together with an additional plasmid encoding Rev or a similar plasmid not coding a functional protein - filled bars. The result show that high titres of vector can be achieved using pESYNGP to supply EIAV Gag/Pol. The highest titres were obtained using the Rev-expressing vector genome plasmid, pONY4.0Z, and they were only slightly lower than observed when Gag/Pol was supplied by pONY3.1. Lower titres were observed with pONY8.0Z vector genome plasmid with pESYNGP than with pONY3.1. This is due to the Rev expression requirement of pONY8.0Z. Rev is expressed by pONY3.1, but not pESYNGP. These results confirm the utility of the codon-optimised Gag/Pol expression plasmid.

Use of the synthetic EIAV gag/pol gene in construction of cell lines which stably express EIAV gag/pol.

[0216] Cells lines which express high amounts of EIAV Gag/pol are required for the construction of packaging and producer cells for EIAV vectors. As a first step in their construction HEK 293 cells were stably transfected with pIRES1hyg ESYNGP (Figure 44 and SEQ ID NO:17), in which EIAV Gag/pol expression is driven by a CMV promoter, and is linked to an ORF for expression of hygromycin phosphotransferase by an EMCV IRES. pIRES1hyg ESYNGP was made as follows. The synthetic EIAV gag/pol gene and flanking sequences was transferred from pESYNGP into pIRES1hygro expression vector (Clontech). First, pESYNGP was digested with *EcoRI*, and the ends filled in by treatment with T4DNA polymerase and then digested with *NotI*. pIRES1hygro was prepared for ligation with this fragment by digestion with *NsiI*, the ends trimmed flush by treatment with T4 DNA polymerase, then digested with *NotI*. Prior to transfection into HEK 293 cells pIRES1hyg ESYNGP was digested with *AhdI* which linearises the plasmid.

[0217] Clonal cell lines, were derived by serial dilution and analysed for expression of Gag/Pol by a Taqman-based product enhanced reverse transcriptase (PERT) assay. Data for the cell line Q3.29, which expressed the highest level of Gag/Pol is shown. The analysis showed that the level of expression from the codon-optimised EIAV Gag/Pol cassette in Q3.29 was very similar to that seen for an EIAV producer line, 8Z.20, in which Gag/Pol is expressed from the pEV53B wild type expression cassette, that produced vector particles at titres of almost 10^6 transducing units per ml. (Figure 45). Assuming exponential amplification during the assay, a difference of Ct value of 1.0 corresponds to a difference of 2-fold in concentration of the reverse transcriptase released from the particles. Therefore the difference in Gag/Pol expression between Q3.29 and 8Z.20 cells is approximately 2-8 fold. Furthermore the Ct values observed indicate that the level of expression of Gag/Pol is significantly higher than in samples of pONY8G vector particles with a titre of 2×10^6 transducing units per ml on D17 cells, but made by transient transfection of HEK 293T cells. These data indicate that the codon-optimised EIAV Gag/Pol construct can be used in the construction of EIAV packaging and producer lines and confirms the previous result that expression is independent of Rev expression.

[0218] The Q3.29 cell line was then tested for its ability to support production of infectious vector particles when transfected with a vector genome plasmid, pONY8.0Z, and the VSV-G envelope expression plasmid, pRV67 and the EIAV REV expression plasmid, pESYNREV. In addition we also evaluated the performance of a plasmid pONY8.3G FB29 (-) which is modified form of the pONY8G vector genome plasmid. PONY8G is a standard EIAV vector genome used for comparison purposes. The modifications and construction of pONY8.3G FB29 (-) (SEQ ID NO:19) are described in PCT/GB00/03837 and briefly are 1) the introduction of loxP recognition sites upstream and downstream of the vector genome cassette 2) the placement of an expression cassette for codon-optimised REV, derived from pESYNREV, and driven by the FB29 U3 promoter downstream of the vector genome cassette and orientated so that the direction of transcription was towards the vector genome cassette. The REV expression cassette is located upstream of the 3' loxP site. Thus the pONY8.3G FB29 - plasmid carries expression cassettes for the vector genome RNA and for EIAV Rev.

[0219] The titres were established by limiting dilution on D17 canine osteosarcoma cells and are shown in Figure 46

[0220] The titres obtained from transfections 2-6 were up to 4.5×10^6 transducing units per ml indicating levels of Gag/Pol expression sufficient to support titres at least this high. The titres obtained were not higher when additional Gag/Pol was supplied (transfection 1) indicating that Gag/Pol expression was not the limitation on titre.

Improved safety profile due to Gag/Pol expression from a codon-optimised expression construct

[0221] RCR formation takes place by recombination between different components of the vector system or by recombination of vector system components with nucleotide sequences present in the producer cells. Although recombination at the DNA level during construction of producer cell lines is possible (perhaps leading to insertional activation of endogenous retroelements or retroviruses) it is thought that recombination to produce RCR occurs mainly between RNA's undergoing reverse transcription, hence occurs within the mature vector particles. In consequence, recombination will be more likely to occur between RNA's which contain packaging signals, such as the vector genome and the *gag/pol* mRNA. Usually however the *gag/pol* transcript is modified so that it is deleted with respect to some or all defined packaging elements, thereby reducing the chances of its involvement in recombination.

[0222] The codon-optimisation process used to create the HIV and EIAV Gag/Pol expression plasmid, pSYNGP and pESYNGP, also results in disruption of sequences and structures that direct packaging as a result of introducing changes at approximately every 3rd nucleotide position. We have obtained evidence for the lower level of incorporation of the codon-optimised RNA derived from pESYNGP into virions.

[0223] The packaging of mRNA's derived from a wild type *gag/pol* pEV53B expression cassette, and from the codon-optimised EIAV *gag/pol* expression cassette, pESYNGP, was compared. Medium was collected from a HEK 293 based cell-lines which were stably transfected with either pEV53B (cell line B-241), or with pESYNGP. Both cell lines produce vector particles which do not contain vector RNA and do not have envelopes. In some experiments, an EIAV vector genome plasmid (pECG3-CZW) was transfected into the cells to serve as an internal positive control for hybridisation and for the presence of particles capable of packaging RNA. pECG3-CZW is a derivative of pEC-LacZ (WO 98/51810) and was made from the latter by 1) reduction of *gag* sequences so that only the first 200nt of *gag*, rather than the first 577nt, was included and 2) inclusion of the woodchuck hepatitis virus post-transcriptional regulatory element (WHV PRE) (corresponding to nt 901-1800 of Acc. No. J04514) into the *NotI* site downstream of the LacZ reporter gene.

[0224] Viral particles derived from each of the cell lines were then partially purified from the medium by equilibrium density gradient centrifugation. To do this 10 ml of medium from producer cells, harvested at 24 hours after induction with sodium butyrate, was layered onto a 20-60% (w/w) sucrose gradient in TNE buffer (pH 7.4) and centrifuged for 24 hours at 25,000 rpm and 4° C in a SW28 rotor. Fractions were collected from the bottom and 10 µl of each fraction assayed for reverse transcriptase activity to locate viral particles. The results of this analysis are shown in (Figure 47) where the profile of reverse transcriptase activity is shown as a function of gradient fraction. In these figures, the top of the gradient is on the right. It should be noted that the levels of RT activity from the pESYNGP-expressing cell were significantly lower than from pEV53B expressing cells. To determine the RNA content of the purified virions, aliquots from the top, middle or bottom fractions were pooled (as indicated by the bars labeled T, M and B) and the RNA from each fraction was subjected to slot-blot hybridization analysis. Using a probe specific for a common region of wild type and synthetic *gag/pol*, encapsidation of RNA was easily detectable in the peak fractions (M) of virions synthesized from the wild type construct (pEV53B), but was not detected from virions synthesized from the synthetic Gag/Pol construct (pESYNGP)(Figure 48). The control for the presence of capsid capable of carrying out encapsidation was the EIAV G3-CZW vector genome which was readily detected in peak fractions from cells expressing either the wild type or synthetic *gag/pol* proteins. Even taking into account the different levels of expression from the wild type and synthetic Gag/Pol expression constructs this result indicates that the RNA from the codon-optimised *gag/pol* gene is packaged significantly less efficiently than the wild type gene and represents a significant improvement to the safety profile of the system. Of further note is that the RNA transcribed from pEV53B was packaged. This RNA is deleted with respect to sequences upstream of the splice donor sequence (CAG/GTAAG) and yet was still packaged. This points to the localisation of major packaging determinants within the *gag* coding region and is in contrast to the collected observations on the location of the packaging signal of HIV-1.

[0225] In additional experiments we have shown that the packaging of transcripts from pEV53B is only slightly lower than from pEV53A (Figure 51). This indicates further that major packaging sequences are located within the *gag* coding region. In these experiments cell line B-241 expressed pEV53B RNA and PEV-17 expressed pEV53A RNA. The EIAV vector genome used to

confirm the presence of packaging competent vector particles was G3-CZR, which is the same as G3-CZW, described above, except for the replacement of the woodchuck post-transcriptional regulatory element with a sequence containing the EIAV RRE elements. Methodology was as described above.

References

[0226]

1. 1. Miller, N., and J. Whelan. 1997. Hum Gene Ther. 8:803-15.
2. 2. Lewis & Emerman. 1993. J. Virol. 68:510.
3. 3. Naldini, L., U. Blomer, P. Gally, D. Ory, R. Mulligan, F. H. Gage, 1. M. Verma, and D. Trono. 1996. Science. 272:263-7.
4. 4. Morgenstern & Land. 1990. Nucleic Acids Res. 18: 3587-3596.
5. 5. Chang, D. D., and P. A. Sharp. 1989. Cell. 59:789-795.
6. 6. Wang & Semenza. 1993. Proc Natl Acad Sci. 90:430.
7. 7. Dachs et al. 1997. Nature Med. 5:515.
8. 8. Firth et al. 1994. Proc Natl Acad Sci. 90: 6496-6500.
9. 9. Madan et al. 1993. Proc Natl Acad Sci. 90:3928.
10. 10. Semenza & Wang. 1992. Mol Cell Biol. 1992. 12: 5447-5454.
11. 11. Takenaka et al. 1989. J Biol Chem. 264: 2363-2367.
12. 12. Peshavaria & Day. 1991. Biochem J. 275: 427-433.
13. 13. Inou et al. 1989. J Biol Chem. 264: 14954-14959.
14. 14. Overell et al. 1988. Mol Cell Biol. 8: 1803-1808.
15. 15. Attenello & Lee. 1984. Science. 226: 187-190.
16. 16. Gazit et al. 1985. Cancer Res. 55: 1660-1663.
17. 17. Yu et al. 1986. Proc Natl Acad Sci. 83: 3194-3198.
18. 18. Dougherty & Temin. 1987. Proc Natl Acad Sci. 84: 1197-1201.
19. 19. Hawley et al. 1987. Proc Natl Acad Sci. 84: 2406-2410.
20. 20. Yee, J. K., A. Miyahara, P. LaPorte, K. Bouic, J. C. Burns, and T. Friedmann. 1994. Proc. Natl. Acad. Sci. USA. 91:9564-8.
21. 21. Jolley et al. 1983. Nucleic Acids Res. 11: 1855-1872.
22. 22. Emerman & Temin. 1984. Cell. 39: 449-467.
23. 23. Herman & Coffin. 1987. Science. 236: 845-848.
24. 24. Bender et al., 1987, J Virol 61: 1639-1646.
25. 25. Pear et al., 1993, Proc Natl Acad Sci 90: 8392-8396.
26. 26. Danos & Mulligan. 1998. Proc Natl Acad Sci. 85: 6460-6464.
27. 27. Markowitz et al. 1988. Virology. 167: 400-406.
28. 28. Cosset et al., 1995, J. Virol. 69: 7430-7436.
29. 29. Mebatsion et al. 1997. Cell. 90: 841-847.
30. 30. Marshall. 1998. Nature Biotechnology. 16: 129.
31. 31. Nature Biotechnology. 1996. 14:556.
32. 32. Wolff & Trubetskoy. 1998. Nature Biotechnology. 16: 421.
33. 33. DuBridge, R. B., P. Tang, H. C. Hsia, P.-M. Leong, J. H. Miller, and M. P. Calos. 1987. Mol. Cell. Biol. 7:379-387.
34. 34. Gey, G. O., W. D. Coffman, and M. T. Kubicek. 1952. Cancer res. 12:264.
35. 35. Kim, S. Y., R. Byrn, J. Groopman, and D. Baltimore. 1989. J. Virol. 63:3708-3713.
36. 36. Adachi, A., H. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. Martin. 1986. J. Virol. 59:284-291.
37. 37. Haas, J., E.-C. Park, and B. Seed. 1996. Current Biology. 6:315.
38. 38. Zolotukhin, S., M. Potter, W. W. Hauswirth, J. Guy, and N. Muzyczka. 1996. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. J Virol. 70:4646-54.
39. 39. Fisher, A., E. Collalti, L. Ratner, R. Gallo, and F. Wong-Staal. 1985. Nature. 316:262-265.
40. 40. Kozak, M. 1992. [Review]. Annu. Rev. Cell Biol. 8:197-225.
41. 41. Cassan, M., N. Delaunay, C. Vaquero, and J. P. Rousset. 1994. J. Virol. 68:1501-8.
42. 42. Parkin, N. T., M. Chamorro, and H. E. Varmus. 1992. J. Virol. 66:5147-51. 68:3888-3895.
43. 43. Mitrophanous K, Yoon S, Rohll J, Patil D, Wilkes F, Kim V, Kingsman S, Kingsman A, Mazarakis N, 1999. Gene Ther. 6 (11): 1808-18
44. 44. Ory, D. S., B. A. Neugeboren, and R. C. Mulligan. 1996. Proc Natl Acad Sci USA. 93:11400-6.
45. 45. Zhu, Z. H., S. S. Chen, and A. S. Huang. 1990. J Acquir Immune Defic Syndr. 3:215-9.
46. 46. Chesebro, B., K. Wehrly, and W. Maury. 1990. J Virol. 64:4553-7.

47. 47. Spector, D. H., E. Wade, D. A. Wright, V. Koval, C. Clark, D. Jaquish, and S. A. Spector. 1990. *J Virol.* 64:2298-2308.
48. 48. Valsesia Wittmann, S., A. Drynda, G. Deleage, M. Aumailley, J. M. Heard, O. Danos, G. Verdier, and F. L. Cosset. 1994. *J Virol.* 68:4609-19.
49. 49. Kim, V. N., K. Mitrophanous, S. M. Kingsman, and K. A. J. 1998. *J Virol* 72: 811-816.
50. 50. Soneoka, Y., P. M. Cannon, E. E. Ramsdale, J. C. Griffiths, G. Romano, S. M. Kingsman, and A. J. Kingsman. 1995. *Nucleic Acids Res.* 23:628-33.
51. 51. Sagerstrom, C., and H. Sive. 1996. RNA blot analysis, p. 83-104. In P. Krieg (ed.), *A laboratory guide to RNA: isolation, analysis and synthesis*, vol. 1. Wiley-Liss Inc., New York.
52. 52. Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen. 1989. *Nature.* 338:254-7.
53. 53. Felber, B. K., M. Hadzopoulou Cladaras, C. Cladaras, T. Copeland, and G. N. Pavlakis. 1989. *Proc. Natl. Acad. Sci. USA.* 86:1495-1499.
54. 54. Hadzopoulou Cladaras, M., B. K. Felber, C. Cladaras, A. Athanassopoulos, A. Tse, and G. N. Pavlakis. 1989. *J. Virol.* 63:1265-74.
55. 55. Schneider, R., M. Campbell, G. Nasioulas, B. K. Felber, and G. N. Pavlakis. 1997. *J Virol.* 71 :4892-903.
56. 56. Schwartz, S., B. K. Felber, and G. N. Pavlakis. 1992. *J. Virol.* 66:150-159.
57. 57. Maldarelli, F., M. A. Martin, and K. Strebel. 1991 *J. Virol.* 65:5732-5743.
58. 58. Mikaelian, I., M. Krieg, M. Gait, and J. Karn. 1996. *J. Mol. Biol.* 257:246-264.
59. 59. Xu, N., C.-Y. Chen, and A.-B. Shyu. 1997. *Mol. Cell. Biol.* 17:4611-4621.
60. 60. Naldini, L. 1998. *Curr. Opin. Biotechnol.* 9:457-463.
61. 61. Parolin, C., T. Dorfman, G. Palu, H. Gottlinger, and J. Sodroski. 1994 *J. Virol.*
62. 62. Dull, T., R. Zufferey, M. Kelly, R. Mandel, M. Nguyen, D. Trono, and L. Naldini. 1998. *J. Virol.* 72:8463-8471.
63. 63. Cui, Y., T. Iwakama, and L.-J. Chang. 1999. *J. Virol.* 73:6171-6176.
64. 64. Chang, L.-J., V. Urlacher, T. Iwakama, Y. Cui, and J. Zucali. 1999. *Gene Ther.* 6:715-728.
65. 65. Harrison, G., G. Miele, E. Hunter, and A. Lever. 1998. *J. Virol.* 72:5886-5896.
66. 66. McBride, M. S., and A. T. Panganiban. 1997. *J. Virol.* 71:2050-8.
67. 67. Lever, A., H. Gottlinger, W. Haseltine, and J. Sodroski. 1989 *J. Virol.* 63:4085-7.
68. 68. Brighty, D., and M. Rosenberg. 1994. *Proc. Natl. Acad. Sci. USA.* 91:8314-8318.
69. 69. Arai, T., M. Takada, M. Ui, and H. Iba. 1999. *Virology.* 260:109-115.
70. 70. Fornerod, M., M. Ohno, M. Yoshida, and I. W. Mattaj. 1997. *Cell.* 90:1051-1060.
71. 71. Fridell, R. A., H. P. Bogerd, and B. R. Cullen. 1996. *Proc. Natl. Acad. Sci. U S A.* 93:4421-4.
72. 72. Pollard, V., and M. Malim. 1998. *Annu. Rev. Microbiol.* 52:491-532.
73. 73. Stade, K., C. S. Ford, C. Guthrie, and K. Weis. 1997. *Cell.* 90:1041-1050.
74. 74. Ullman, K. S., M. Powers, A" and D. J. Forbes. 1997. *Cell.* 90:967-970.
75. 75. Otero, G. C., M. E. Harris, J. E. Donello, and T. J. Hope. 1998. *J. Virol.* 72:7593-7597.
76. 76. Wolff et al. 1997. *Chem Biol.* 4: 139-147.

SEQUENCE LISTING

[0227]

<110> Oxford Biomedica (UK) Limited

<120> Method

<130> P008827EPA

<150> GB 0009760.0

<151> 2000-04-19

<160> 37

<170> Patentln version 3.5

<210> 1

<211> 4307

<212> DNA

<213> Human immunodeficiency virus

<400> 1

```

atgggtgcga gaggctcagt attaagcggg ggagaattag atcgatggga aaaaattcgg 60
ttaagccag ggggaagaa aaaatataaa ttaaaacata tagtatgggc aagcagggag 120
ctagaacgat tccaggttaa tcttggcctg ttagaacatc cagaaggctg tagacaaata 180
ctgggacagc tacaaccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtacgaa cctctcattg tgtgcatcaa aggatagaga taaagacac caaggagct 300
ttagacaaga tagaggaaga gcaaaacaaa agtaagaaaa agccacagca agcagcagct 360
gacacaggac acagcaatca ggtcagccaa aattaccccta tagtcagaaa catccagggg 420
caaatggtac atcaggccat atcacctaga actttaaatg catgggtaaa agtagtagaa 480
gagaaggctt tcagcccaga agtgataccc atgttttcag cattatcaga aggagccacc 540
ccacaagatt taaacccat gctaaacaca gtgggggggc atcaagcagc catgcaaatg 600
ttaaaagaga ccatcaatga ggaagctgca gaattgggata gagtgcaccc agtgcatgca 660
gggcctattg caccaggcca gatgagagaa ccaaggggaa gtgacatagc aggaactact 720
agtagccctc aggaacaaat aggatggatg acaataaata cacctatccc agtaggagaa 780
attttataaaa gatggataat cctgggatta aataaaatag taagaatgta tagccctacc 840
agcattctgg acataagaca aggaccaaag gaaccttta gagacta-gt agacgggttc 900
tataaaactc taagagccga gcaagcttca caggaggtaa aaaattggat gacagaaacc 960
tgtttggtcc aaaatcgcaa cccagattgt aagactattt taaaagcatt gggaccagcg 1020
gctacactag aagaatggt gacagcatgt caggagtag gaggacccgg ccataaggca 1080
agagtttttg ctgaagcaat gagccaaagta acaaatccag ctaccataat gatgcagaga 1140
ggcaatttta ggaaccaaa aaagattgtt aagtgtttca attgtggcaa agaagggcac 1200
acaggcagaa attgcagggc ccttaggaaa aagggtctgt ggaatgtgtg aaaggaagga 1260
caccaaatga aagattgtac tgagagacag gctaattttt tagggaaagt ctggccttcc 1320
tacaagggaa ggccagggaa ttttcttcag agcagaccgc agcccaacagc cccaccagaa 1380
gagagcttca ggtctgggt agagacaaca actccccctc agaagcagga gccgatagac 1440
aaggaactgt atcctttaac tccccctcag tcaactcttg gcaacgacc ctcgtcacaa 1500
taaagatagg ggggcaacta aaggaaagctc tattagatac aggcagagat gatcacgat 1560
tagaagaatg gagtttgcca ggaagatgga aaccaaaat gataggggga attggaggtt 1620
ttatcaaatg aagacagtat gatcagatac tcatagaaat ctgtggacat aaagctatag 1680
gtacagtttt agtaggacct acacctgtca acataattgg aagaatctgt ttgactcaga 1740
ttgtttgcac ttttaatttt cccattagcc ctattgagac tgtaccagta aaattaaagc 1800
caggaatgga tggcccaaaa gttaaacaat ggccattgac agaagaaaaa ataaaagcat 1860
tagtagaaat ttgtacagag atggaaaagg aagggaatat ttcaaaaatt gggcctgaaa 1920
atccatacaa tactccagta tttgccataa agaaaaaaga cagtactaaa tggagaaaat 1980
tagtagattt cagagaactt aataagagaa ctcaagactt ctgggaagt ctcaattaggaa 2040
taccacatcc cgcagggtta aaaaagaaaa aatcagtaac agtactggat gtgggtgatg 2100
catatttttc agttccctta gatgaagact tcaggaaagta tactgcattt accataccta 2160
gtataaacaa tgagacacca gggattagat atcagtacaa tgtgcttcca cagggatgga 2220
aaggatcccc agcaattatc caaagtagca tgacaaaaat cttagagcct tttagaaaaa 2280
aaaaatccaga catagtattc tatcaatata tggatgattt gtatgtagga tctgacttag 2340
aaatagggca gcatagaaca aaaatagagg agctgagaca acatctgttg aggtggggac 2400
ttaccacacc agacaaaaaa catcagaag catctccatt cctttggatg ggttatgaac 2460
tccatcctga taaatggaca gtacagccta tagtgctgcc agaaaaagac agctggactg 2520
tcaatgacat acagaagtta gtggggaaat tgaattgggc aagtcagatt taccagggga 2580

```

```

ttaaagtaag gcaattatgt aaactcctta gaggaaccaa agcactaaca gaagtaatac 2640
cactaacaga agaagcagag ctagaactgg cagaaaacag agagattcta aaagaaccag 2700
tacaatggag gtattatgat ccatcaaaag acttaatagc agaatatacag aagcaggggc 2760
aaggccaatg gacatatcaa atttatcaag agccatttaa aaatctgaaa acaggaaaaa 2820
atgcaagaat gaggggtgac cacactaatg atgtaaaaa attaacagag gcagtgcaaa 2880
aaataccac agaaaagcata gtaatatggg gaaagactcc taaatttaaa ctgcccatac 2940
aaaaggaaac atgggaaaca tgggtggacag agtattggca agccaccgg attcctgagt 3000
gggagtttgt taataccctt ccttagtgta aattatggta ccagtttaga aaagaaccca 3060
tagtagggag agaaaccttc tatgtagatg gggcagctaa caggagagct aaattaggaa 3120
aagcaggata tgttactaat agaggaagac aaaaagtgt caccctaact gacacaacaa 3180
atcagaagac tgagttacaa gcaatttata tagctttgca ggattcggga ttagaagtaa 3240
acatagtaac agactcacaa tatgcattag gaatcatcca agcacacca gatcaaatg 3300
aatcagagtt agtcaatcaa ataataagag agttaataaa aaaggaaaag gtcctatctg 3360
catgggtacc agcacacaaa ggaattggag gaaatgaaca agtagataaa ttagttagtg 3420
ctggaatcag gaaagtacta tttttagatg gaatagataa ggcccaagat gaacatgaga 3480
aatatcacag taattggaga gcaatggcta gtgattttaa cctgccacct gtagtagcaa 3540
aagaatatgt agccagctgt gataaatgtc agctaaaagg agaagccatg catggacaag 3600
tagactgtag tccaggaata tggcaactag attgtacaca tttagaagga aaagtatatc 3660
tggtagcagt tcatgtagcc agtggatata tagaagcaga agttattcca gcagaacag 3720
ggcaggaaac agcatatttt cttttaazat tagcaggaag atggccagta aaaacaatac 3780
atactgacaa tggcagcaat ttcacccgtg ctacggttag ggccgcctgt tggtagggcg 3840
gaatcaagca ggaatttggg attccctaca atccccaaag tcaaggagta gtagaatcta 3900
tgaataaaga attaaagaaa attataggac aggttaagaga tcaggctgaa catcttaaga 3960
cagcagtaca aatggcagta ttcatccaca attttaaaag aaaagggggg attggggggt 4020
acagtgcagg ggaagaata gttagacata tagcaacaga catacaaact aaagaattac 4080
aaaaacaaat tacaaaaatt caaaattttc gggtttatta caggacagc agaaattcac 4140
tttggaaagg accagcaaa cctcctctgga aaggtgaagg ggcagtagta atacaagata 4200
atagtacat aaaagttagt ccaagaagaa aagcaagat cattagggat tatggaaaac 4260
agatggcagg tgatgattgt gtggcaagta gacaggatga ggattag 4307

```

<210> 2

<211> 9772

<212> DNA

<213> Artificial Sequence

<220>

<223> pSYNGP

<400> 2

```

tcaatattgg ccattagcca tattattcat tggttatata gcataaaaca atattggcta 60
ttggccattg ccatcgtttg atctatatca taatatgtac atttataatg gctcatgtcc 120
aatatgacgc ccatgtttgg attgatttat gactagttaa taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taataggccc 240
gcctggctga ccgcccaacg acccccgccc attgacgtca ataatacgt atgttcccat 300
agtaacgcca atagggaact tccattgacg tcaatgggtg gagtatttac ggtaaaactg 360
ccacttggca gtacatcaag tgtatcatat gccaaagtcc cccctaatg acgtcaatga 420
cggtaaatgg cccgcctggc attatgcccc gtacatgacc ttacgggact ttctacttgg 480
gcagtcacac tacgtattag tcatcgctat taccatgggt atgcggtttt ggcagtacac 540
caatgggcgt ggatagcggg ttgactcccg gggatttcca agtctccaco ccattgacgt 600
caatgggagt ttgttttggc accaaaatca accgggacttt ccaaaatgto gtaacaactg 660
cgatcccgcc ccccgttgac gcaaatgggg ggttaggcgtg tacgggtggga ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780
agttaaatgg ctaacgcagt cagtgttctt gacacaacag tctcgaaact aagctgcagt 840
gactctotta aggtagcctt gcagaagtgt gtctgaggg actgggcagg taagtatcaa 900
ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttc tgatagggac ctatttggct tactgacatc cactltgccl ttctctccac 1020
agggtgccac tcccagttca attacagctc tttaaggctag agtacttaat acgactcact 1080
ataggctagc ctgcagaatt ccgccacatg ggcgcgcgcg ccagcgtgct gtcgggcgcg 1140
gagctggacc gctggcgaga gatccgctg ccgcccgcg gcaaaaagaa gtacaagctg 1200
aagcacatcg tgtgggcccg ccgcgaactg gagcgcttcg ccgtgaaccc cgggctcctg 1260
gagaccagcg aggggtgccc ccagatccct ggccaaactg agccagcct gcaaaccgcg 1320
agcgaggagc tgcgcagcct gtacaacacc gtggccacgc tgtactgctg ccaccagcgc 1380
atcgaaatca aggatacgaa agaggccctg gataaaatcg aagaggaaca gaataagagc 1440

```

```

aaaaagaagg cccaacaggc cgcgcgggac accggacaca gcaaccaggt cagccagaac 1500
tccccatcg tgcagaacat ccaggggcag atgggtgcacc agggcacttc ccccgccagc 1560
ctgaacgcct ggtggaaggt ggtggaagag aaggcttcta gcccgagggt gataccatg 1620
ttctcagccc tgcagagggg agccaccccc caagatctga acacatgct caacacagt 1680
gggggacacc agcccgccat gcagatgctg aaggagacca tcaatgagga ggtgcccga 1740
tgggattcgt tcatcccggt gacgcagagg cccatcgcac cgggccagat gctgagacca 1800
cggggctcag acatcgccgg aacgactagt acccttcagg aacagatcgg ctggatgacc 1860
aacaacccc acatcccggt gggagaaatc tacaacgct ggatcatcct gggcctgaac 1920
aagatcgtgc ccatgtatga cctaccagc atcctggaca tccgccagg cccgaaggaa 1980
cccttccggt actcgttga ccggttctac aaaaacgctc ggcgcagaca ggtlagccag 2040
gaggtgaaga actggatgac cgaacccctg ctggtccaga acgcgaaccc ggaactgcaa 2100
acgatcctga agccctcagg ccacgcgctc accctagagg aaatgatgac cgcctgtcag 2160
ggagtgggcg gacccggcca caaggcacgc gtccctggctg agggcactgag ccaggtgac 2220
aactccgcta ccatcatgat gcagcgccgg aacttccgga accaacgcaa gatcgtcaag 2280
tgcctcaact gtggcaaaag agggcacaca gcccccaact gcagggcccc taggaaaaag 2340
ggctgttggg aatgtggaaa ggaaggacac caaatgaaag attgtactga gagacaggct 2400
aattttttag ggaagatcgt ccttccccc aagggaaagg cagggaattt tcttcagagc 2460
agaccagagc caacagcccc accagaagag agcttcaggt ttggggaaga gacaacaact 2520
ccctctcaga agcaggagcc gatagacaag gaactgtatc cttagcttc cctcagatca 2580
ctctttggga gcgacccctc gtcacaataa agataggggg gcagctcaag gaggctctcc 2640
tggacacggg agcagaagac accgtgctgg aggagatgct gttgccagg cgttggaagc 2700
cgaagatgat cgggggaatc ggcggttcca tcaaggtgcg ccagtatgac cagatcctca 2760
tcgaatatcg cggccacaag gctatcggtg ccgtgctggt gggcccccaca cccgtcaaca 2820
tcatcgagcg caactcgttg acgcagatcg gttgcacgct gaacttcccc attagcccta 2880
tcgaagcgtg accgtggaag ctgaagcccg ggaatggagg ccggaaggtc aagcaatggc 2940
cattgacaga ggagaagatc aaggcaactg tggagatttg cacagagatg gaaaaggaa 3000
ggaaaatcct caagatttgg cctgagaacc cgtacaacac gcggtgttc gcaatcaaga 3060
aagaagactc gacgaaatgg cgaagctgg tggacttccg cgagctgaac aagcgcacgc 3120
aagaactcgt ggaagttcag ctgggcatcc cgcaccccg agggctgaag aagaagaaat 3180
ccgtgaccgt actggatgtg ggtgatgctc acttctccgt tccctggac gaagacttca 3240
ggaagtacac tgccttcaca atcccttoga tcaacaacga gacaccggg attcgatca 3300
agtacaactg gctgcccag gctggaaag gctctcccg aatcttcag agtagcatga 3360
ccaaaatcct ggagccttcc cgcacaacga accccgacat cgtcatctat cagtacatg 3420
atgacttcta cgtgggtctc gatctagaga tagggcagca ccgcaccaag atcgaggag 3480
tgcccgacga cctgttgagg tgggactga ccacaccga caagaagcac cagaaggagc 3540
ctcccttctc tggatgggtg tacgagctgc accctgacaa atggaccgtg cagcctatcg 3600
tgcctccaga gaagacagc tggactgtca acgcacata gaagctggtg ggggaagtga 3660
actggccagc caagatttac ccagggatta aggtgaggca gctgtgcaaa ctctcccg 3720
gaaccaaggc actcacagag gtgatccccc taaccagga ggcagagctc gaactggcag 3780
aaaaacgaga gacctaagag gagcccgctg accgctgta ctatgacccc tccaaggacc 3840
tgatcccgga gatccagaag caggggcaag gccagtgag cctatcagatt taccaggagc 3900
ccttcaagaa cctgaagacc ggcaagtacg cccggatgag ggggtcccac actaacga 3960
tcaagcagct gaccgaggcc gtgcagaaga tcaccacga aagcatcgtg atctgggaa 4020
agactcctaa gttcaagctg cccatccaga aggaacactg ggaacactgg tggacagagt 4080
attggcagcg cactgtgatt cctgagtggt agttcgtcaa caccctctcc ctggtgaagc 4140
tgtgttacca gctggagaag gagcccatag tgggcgcgga aacttctac tgggatggg 4200
ccgctaacag ggagactaag ctgggcaaa gcgatacgt cactaacggg ggcagacaga 4260
aggttgtcac cctcactgac accaccaaacc agaagactga gctgcaggcc atttacctg 4320
ctttgcagga cctggcgctg gaggtgaaca tctgtacaga cctcagtat gccctgggca 4380
tcatcaagc ccagccagac cagagtgaat ccgagctggt caatcagatc atcgagcagc 4440
tgatcaagaa ggaaaaggtc tatctggccl ggttaccgc ccaaaaaggc attggcgga 4500
atgagcaggt cgacaagctg gtctcgctg gcatcaggaa ggtgctattc ctggatggca 4560
tcgacaagc ccaggcagag caccgagaat accacagcaa ctggcgggcc atggctagc 4620
actcaccct gcccccctg gtggccaaag agatcgtggc cagctgtgac aagtgtagc 4680
tcaaggcgga agccatgcat ggccagggtg actgtagccc cggcatctgg caactcgatt 4740
gcacccatct ggaggcgag gttatcctgg tagccgtcca tgtggccagt ggtacatcg 4800
agggcgaggt catttccgcc yaaacagggc ctacttctc ctgaagctg 4860
caggccgggt ccagtgaa gaccatccata ctgacaatg cagcaatttc accagtcta 4920
cggttaaag ccgctgctg tggcgggaa tcaagcagga gttcgggata ccttacaatc 4980
cccagagtca gggcgtgctc gagtctatga ataaggagtt atcgcccaag 5040
tcagagatca ggttagcat ctcaagacc cgttccaaat ggcggttatc atccacaatt 5100
tcaagcgaa ggggggatt ggggggtaca gtgcggggga cggatctg gacatcatg 5160
cgaccagat ccagactaag gagctgcaaa agcagattac caagattcag aatttccgg 5220
tctactacag ggacagcaga aatccctct ggaaggccc agcgaagctc ctctggaag 5280

```

```

gtgagggggc agtagtgatc caggataata gcgacatcaa ggtggtgccc agaagaaagg 5340
cgaagatcat tagggattat ggcaaacaga tggcggtgga tgattgctgt gcgagcagac 5400
aggatgagga ttaggaaattg ggctagagcg gcgcgttccc tttagtgagg gttaatgctt 5460
cgagcagaca tgataagata caatgatgag tttagacaaa ccacaactag aatgcagtga 5520
aaaaaatgct ttatttgga aatttgatgat gctattgctt tatttgtaac cattataago 5580
tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt tcagggggag 5640
atgtgggagg ttttttaag caagtaaaac ctctacaaat gtggtaaaat cagataagga 5700
tcgatccggg ctggcgtaat agcgaagagg ccgcaccga tgcctctcc caacagttgc 5760
gcagcctgaa tggcgaattg acgcgccctg tagcgggcca ttaagcgagg cgggtgttgt 5820
ggttacgcgc agcgtgacgc ctacacttgc cagcgcccta gcgcgcctc ctctcgcttt 5880
cttcccttcc ttctcgcca cgttcgccgg ctctcccggt caagctctaa atcgggggct 5940
cccttttagg ttccgattta ggcctttaag gcacctcgac cgcataaac ttgatttggg 6000
tgatggttca cgtagtgggc catcgccctg atagacggtt ttctgcctt tgacgttga 6060
gtccacgttc tttaattagt gactctgtgt ccaactgga acaactca accctatctc 6120
ggtctattct ttgatattat aagggaattt gccgatttc gccatttgt taataaatga 6180
gctgatttaa caaatattta acggaattt taacaaaata ttaacgttta caatttcgc 6240
tgatcggtga tttctcctt acgcattctg gcggtatttc acacgcata cgcggtctg 6300
cgacgaccca tggcctgaaa taacctctga aagaggaaat tgggttaggt cctctctagg 6360
cggaagaaac cagctgtgga atgtgtgtca gttagggtgt gaaagtccc caggctccc 6420
agcaggcaga agtatgcaa gcacatctct caattagtca gcaaccaggt gtggaaagt 6480
cccaggctcc ccaggcgga caagtatgca aagcatgcat ctcaattagt cagcaaccat 6540
agtccgcgcc ctaactccgc ctaactccgc ccaactccgc cccagttccg ccaattctcc 6600
gcccatggc tgactaattt tttttattta tgcagaggcc gaggccgctt cggcctctga 6660
gctattccag agtagtgag gaggcttttt tggaggccca ggcttttgca aaaagcttga 6720
ttctctgac acaacagctc ogaacttaag gctagagcca ccattgttga acaagatga 6780
ttgcacgcag gttctccggc cgtctgggtg gagaggtcat tggctatga ctgggcacaa 6840
cagacaatgc gctgctctga tgcgcgctg ttcggctgt cagcgaggg gcgcgcgtt 6900
ctttttgtca agaccgacct gtcgctgccc ctgaatgaac tgcaggacga ggcagcgagg 6960
ctatcgtggc tggccacgac gggcgttccc tgcgcagctg tgcctgagct tgcactgaa 7020
gcgggaaggg actggtgct atttggcga gtcgcggggc aggatctccc gtcactccac 7080
cttgcctccc cggagaagg atccatcatg gctgatgcaa tgcggcggtt gcatcagctt 7140
gatccggcta cctgcaccat cgaaccacaa gcgaacatc gcactcgagg agcacgtact 7200
cggatggaag cgggtcttgt cgaatcaggat gatctggaag aagagcatca ggggtctgag 7260
ccagcgaac tgttcggcag gctcaaggcg cgcctgcccg acggcgagga tctcgtctg 7320
accctggcg atgctctgct tgcgaatata atgggtgaaa atggcgctt tctcggattc 7380
atcgactgtg gccggtggg tgtggcggac cgtatcagg acatagcgtt ggtcaccgtt 7440
gatattgct aagaycttgg cgtgaccgct gcctgctgt lctcgtgtt ltaaggtatc 7500
gccgctccc attcgacgc catcgccctt tatcgccctt ttgacgagtt ctctgtagcg 7560
ggactctggg gttcgaatat accgaaccaag cgaagcccaa cctgccatca cgtatggcgc 7620
aataaaatat ctttattttt attacatctg tgtgttggtt tttgtgtga atcgatagcg 7680
ataaagatcc cgtatgtgtg cactctcagt acaatctgct ctgatgcgcg atagttaaagc 7740
cagcccccgc acccgccaac acccgctgac gcgcctgac gggcttgtct gctcccggca 7800
tcgccttaca gacaagctgt gaccgtctcc gggagctgca tgtgtcagag gttttcaccg 7860
tcaccccgga aacgcgcgag acgaaagggc ctctgataac gctattttt ataggttaat 7920
ghcatgataa taattggttc ttagaactca ggtggcaatt ttccgggaaa tgtgcgcgga 7980
accctatttt gtttattttt ctaaatatct tcaaatatgt atccgctcat gagacaataa 8040
ccctgtaaaa tgcctcaata atattgaaaa aggaagagta tgaatttca acatttcgt 8100
gtcgccctta ttcccttttt tgcggcattt tgccttctgt tttttgctca cccgaaacg 8160
ctggtagaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta catcgaactg 8220
gatctcaaca cgggtaagat ccttgagagt ttgcgcgcgc aagaacgttt tccaatgatg 8280
agcactttta aagtctctgt atgtggcgcg gtattatccc gtattgcgc cgggcaagag 8340
caactcggct gcgcataca ctattctcag aatgaacttg ttgagtactc accagtcaca 8400
gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgtctc cataaccatg 8460
agtgaataca ctcgggccaa cttacttctg acaacgatcg gaggaccgaa gggactaac 8520
gcttttttgc acaacatggg ggatcatgta actegccttg atcgttggga accggagctg 8580
aatgaagcca taccaaacga cgaagctgac accacgatgc ctgtagcaat ggcaacacg 8640
ttgcgcaaac tattactgg cgaactactt actctagctt cccggcaaca attaatagac 8700
tggatggagg cggataaagl tgcaggacca ctctctgctt cggcccllcc ggttggctgy 8760
tttatgtctg ataaatctgg agcgggtgag cgtgggtctc gcggtatcat tgcagcactg 8820
gggcccagatg gtaagccctc ccgtatcgta gttatctaca cgaaggggag tccaggcaact 8880
atggatgaac gaaatagaca gatcgtgag ataggtgctt cactgattaa gcattggtaa 8940
ctgtcagacc aagtttaact atalatactt tagattgatt taaaacttca tttttaattt 9000
aaaaggatct aggtgaagat cctttttgat aatctcatga ccaaatccc ttaactgtag 9060
ttttcgttcc actgagcgct agaccccgta gaaaagatca aaggatcttc ttgagatctt 9120

```

```

ttttttctgc gogtaactct ctgcttgcaa acaaaaaaac caccgctacc agcgggtggtt 9180
tgtttgccgg atcaagagct accaactctt ttccgaagg taactgggtt cagcagagcg 9240
cagataccaa atactgtcct tctagttag ccgtagttag gccaccaatt caagaactct 9300
gtagcacccg ctacatacct cgctctgcta atcctgttac cagtggctgc tgcagtgcc 9360
gataagtcgt gtcttaccgg gttgactca agacgatagt taccggataa ggcgcagcgg 9420
tcgggctgaa cggggggttc gtgcacacag ccagcttgg agcgaacgac ctacacgaa 9480
ctgagatacc tacagcgtga gctatgagaa agcgcacgac ttccgaagg gagaagggcg 9540
gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga gcttccaggg 9600
ggaacgcctt ggtatcttta tagtctgtc gggtttcgac acctctgact tgagcgtcga 9660
ttttttgtag gctcgtcagg ggggcggagc ctatggaaaa acgcagacaa cgcggcctt 9720
ttacggttcc tggccttttg ctggcctttt gctcacatgg ctgcacagat ct 9772

```

<210> 3

<211> 2571

<212> DNA

<213> Human immunodeficiency virus

<400> 3

```

atgagagtga aggggatcag gaggaattat cagcactggt ggggatgggg cacgatgctc 60
cttgggttat taatgatctg tagtgctaca gaaaaattgt gggtcacagt ctattatggg 120
gtacctgtgt ggaagaagcg aaccaccact ctattttgtg catcagatgc taaagcatat 180
gatacagagg tacataatgt ttgggccaca caagcctgtg taccacacaga cccaaccca 240
caagaagtga aatttgtaaa tgtgacagaa aattttaaca tgtggaaaaa taacatggta 300
gaacagatgc atgaggatat aatcagttta tgggatcaaa gcctaaagcc atgtgtaaaa 360
ttaacccac tctgtgttac tttaaattgc actgatttga ggaataacac taataccaat 420
aatagtactg ctaataacaa tagtaatagc gaggaacaa taaagggagg agaattgaaa 480
aactgctctt tcaatatcac cacaagcata agagataaga tgcagaaga atalgcactt 540
ctttataaac ttgataatgt atcaatagat aatgatagta ccagctatag gttgataagt 600
tgtataacct cagtcattac acaagcttgt ccaagatat cctttgagcc aattcccata 660
cactatttgt ccccgctggg ttttgcgatt ctaaaatgta acgataaaaa gttcagtggg 720
aaaggatcat gtaaaaatgt cagcacagta caatgtacac atggaattag gccagtagta 780
tcaactcaac tgcgtgttaa tggcagtcct gcagaagaag aggtagtaat tagatctgag 840
aatttcactg ataatgctaa aaccatcata gtacatctga atgaatctgt acaaattaat 900
tgtacaagac ccaactacaa taaaagaaaa aggatacata taggaccagg gagagcattt 960
tataacacaa aaaatataat aggaactata agacaagcac attgtaacat tagtagagca 1020
aaatggaatg acactttaag acagatagtt agcaaatlaa aagaacaatt taagaataaa 1080
acaatagctt ttaalcaatc ctccaggagg gacccagaaa ttglaatgca cagtttlaal 1140
tgtggagggg aatttttcta ctgtaataca tcaccactgt ttaatagtag ttggaatggg 1200
aataataactt ggaataatac tacaggggtc aataacaata tcacacttca atgcaaaaa 1260
aaacaaatta taacacatgt gcaggaagta ggaaaagcaa tgtatgcctc tccattgaa 1320
ggacaaatta gatgttctac aatatattaca gggctactat taacaagaga tgggtgtaag 1380
gacacggaca cgaacgacac cgagatcttc agacctggag gaggagatat gagggacaat 1440
tgggaagtg aattatataa atataaagta gtaacaattg aaccattagg agtagcacc 1500
accagggcaa agagaagagt ggtgcagaga gaaaaaagag cagcgatagg agctctgttc 1560
cttgggttct taggagcagc aggaagcact atggcgcgag cgtcagtgac gctgacggta 1620
caggccagac tattattgtc tggatatagt caacagcaga acaatttgct gagggccatt 1680
gaggcgcaac agcatatgtt gcaactcaca gtctggggca tcaagcagct ccaggcaaga 1740
gtctcggctg tggaaagata cctaaggat caacagctcc tgggcttttg gggttgctct 1800
ggaaaactca tttgcaccac tactgtgcct tggaaatgcta gttggagtaa taaatctctg 1860
gatgatattt gcaataacat gacctggatg cagtgggaaa gagaatttga caattacaca 1920
agctlaatat actcattact agaaaaatcg caaacccaac aagaaaaaga tgaacaagaa 1980
ttattggaat tggataaatg ggcaagtttg tggaaattgg ttgacataac aaattggctg 2040
tggtatataa aaatattcat aatgatagta ggaggcttgg taggtttaag aatagttttt 2100
gctgtacttt ctatagttaa tagagttagg cagggaact caccattgtc gttgcagacc 2160
cgccccccag ttccgagggg acccgacagg cccgaaggaa tcgaagaaga aggtggagag 2220
agagacagag acacatccgg tcgattagtg catggattct tagcaattat ctgggtcgac 2280
ctgcgagacc tgttcctctt cagctaccac cacagagact tactcttgat tgcagcgagg 2340
attgtggaac ttctgggaag caggggttgg gaaagtctca aatattgggt gaatctccta 2400
cagttattga gtcaggaact aaagagtagt gctgttagct tgtttaatgc cacagctata 2460
gcagtagctg aggggacaga tagggttata gaagtactgc aaagagctgg tagagctatt 2520
ctccacatac ctacaagaat aagacagggc ttggaaaggg ctttgctata a 2571

```

<210> 4

<211> 2571

<212> DNA

<213> Artificial Sequence

<220>

<223> SYNp-160mn - codon optimised env sequence

<400> 4

```

atgagggtga aggggatccg ccgcaactac cagcactggt ggggctgggg cacgatgttc 60
ctggggctgc tgaatgatctg cagcgccacc gagaagctgt gggtgaccgt gtactacggc 120
gtgcccggtg ggaaggaggc caccaccacc ctgttctgcg ccagcgacgc caaggcgtag 180
gacaccgagg tgcacaacgt gtggccacc caggcgtagc tgcccaccga cccaacccc 240
caggaggtgg agctcgtgaa cgtgaccgag aacttcaaca tgtggaagaa caaatggtg 300
gagcagatgc atgaggacat catcagcctg tgggaccaga gcctgaagcc ctgctgaaag 360
ctgacccccc tgtgcgtgac cctgaactgc accgacctga ggaacaccac caacaccaac 420
aacagcaccc ccaacaacaa cagcaacagc gagggcacca tcaaggcgcg cgagatgaag 480
aactgcagct tcaacatcac caccagcctc cgcgacaaga tgcagaagga gtacgcccgt 540
ctgtacaagc tggatatcgt gagcatcgac aacgacagca ccagctaccg cctgatctcc 600
tgcaacaacca gctgatcac ccaggcctgc cccaagatca gcttcgagcc catcccatc 660
cactactcgc ccccgccggg cttcgccatc ctgaagtga acgacaagaa gttcagcggc 720
aagggcagct gcaagaacgt gagcaccgtg cagtgcaccc acggaatccg gcggtggtg 780
agcaccagc fcctgctgaa cggcagcctg gccgaggagg aggtggtgat ccgagcgag 840
aacctccccc acaacgccaa gaccatcac gtgcacctga atgagagcgt gcagatcaac 900
tgcacgcgtc ccaactacaa caagcgcaag cgcattccca tcggccccgg gcgcgcttc 960
tacaccacca agaacatcat cggcaccatc cgcaggccc actgcaacat ctctagagcc 1020
aagtggaaag acaccctgcg ccagatcgtg agcaagctga agggagcagt caagaacaa 1080
accatcgtgt tcaaccagag cagcgccggc gaccccgaga tctgtatgca cagcttcaac 1140
tgcgcgcggt aattcttcta ctgcaacacc agccccctgt tcaacagcac ctggaacggc 1200
aacaaacact ggaacaacac caccggcagc aacaacaata ttacctcca gtgcaagatc 1260
aagcagatca tcaacatgtg gcaggagggt ggcaaggcca tgtacgcccc ccccatcgag 1320
ggccagatcc ggtgcagcag caacatcacc ggtctgctgc tgaccgcga cggcgccaag 1380
gacaccgaca ccaacgacac cgaatcttc cgcgccggcg gcggcgacat gcgcgacaac 1440
tggagatctg agctgtacaa gtacaagggt gtgacgatcg agccccctgg cgtggcccc 1500
accaaggcca agcgcgcggt ggtgcagcgc gagaagcggg ccgcatcgg cgcctgttc 1560
ctgggcttcc tggggcgcg cggcagcacc atggggcgcg ccagcgtgac cctgacgtg 1620
caggcccgcc tgcctctgag cggcatcgtg cagcagcaga acaacctcct ccgcccctc 1680
gaggccagc agcatatgct ccagctcacc gtgtggggca tcaagcagct ccaggcccg 1740
gtgctggcgg tggagcgcta cctgaaggac cagcagctcc tgggctctgt gggctgctcc 1800
ggcaagctga tctgcaccac caccgtaccc tggaaacgct cctggagcaa caagagcctg 1860
gacgacatct ggaacaacat gacctggatg cagtgggagc gcgagatcga taactacac 1920
agcctgatct acagcctgct ggagaagagc cagaccagc agggagaaga cgagcaggag 1980
ctgctggagc tggacaagtg ggcgagcctg tggaaactgt togacatcac caactggctg 2040
tggatcatca aatcttcat catgattgtg ggcgcctgg tgggcctcgg catcgtgttc 2100
gccgtgctga gcatcgtgaa ccgcgtgcgc cagggtatca gccccctgag cctccagacc 2160
cggccccccc tgcgcgcgg gcccgaccgc ccgaggggca tccgaggaga gggcgcgag 2220
cgcgaccggg acaccagcgg caggctcgtg caccgcttcc tggcgatcat ctgggtcgac 2280
ctccgcagcc tgttcctgtt cagctaccac caccgagacc tgctgctgat cgcgcgccgc 2340
atcgtggaa ccttagggcg ccgcggtgg gaggtggtga agtactggtg gaacctctc 2400
cagtatggg gccaggagct gaagtccagc gccgtgagcc tgcgaaacgc caccgcatc 2460
gcggtggcg agggcaccca ccgctgata gaggtgctcc agaggcgcg gagggcgatc 2520
ctgcacatcc ccaccgcat ccgcagggg ctcgagaggg cgtgctgta a 2571

```

<210> 5

<211> 10112

<212> DNA

<213> Artificial Sequence

<220>

<223> pESYNGP

<400> 5

tcaatattgg ccattagcca tattattcat tggttalata gcataaatca atattggcta 60

```

ttggccattg cacaagtgtg atctatatca taaatgtgac atttatattg gctcatgtcc 120
aatatgaccg ccaatgltggc altgattatt gactagttaa taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taatgggccc 240
gcctggcgtg ccgcccacag acccccgcgc attgacgtca ataagacgt atgttcccat 300
agtaaaccca atagggagct tccattgacg tcaatgggtg gattatttac ggtaaaactg 360
ccacttggca gtacatcaag tgtatcatat gccaaagtcc cccctattg acgtcaatga 420
cggtaaatgg ccgcctgggc attatgcccc gtacatgacc ttacgggagc ttctactctg 480
gcagtacatc tacgtattag tcatcgctat taccatgggt atgcgggttt ggcagtacac 540
caatgggggt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatggggag ttgttttggc accaaaatca accgggacttt ccaaaatgtc gtaacaaactg 660
cgatcgcccg ccccgttgac gcaaatgggc ggtaggcgtg tacgggtggg ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgctgg agtttatcac 780
agttaaattg ctaacgcagt cagtgttctc gacacaacag tctcgaactt aagctgcagt 840
gactctotta aggtagcctt gcagaagtgt gtogtgaggc actgggcagg taagtatcaa 900
ggttacaaga caggtttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttc tgataggcac ctattggtct tactgacatc cactttgctt ttctctccac 1020
aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaat acgactcact 1080
atagggtaga gaattcgcca ccatgggcga tccctcacc tgggtccaaag ccoctgaagaa 1140
actggaaaaa gtcaccgttc aggttagcca aaagcttacc acaggcaatt gcaactgggc 1200
attgtccctg gtggatcttt tccacgacac taatttcgtt aaggagaaag attggcaact 1260
cagagccgtg atccccctct tggaggagct gaccnaaaat ttgtctgggc agggagcgca 1320
agctttcgag ccacctgggt gggccatcag ccagctcaaa atggggctgc aaatcaacaa 1380
cgtggttgac ggttaagctg gctttcaact gctcgcgct aagtacgaga agaaaaacgc 1440
caacaagaaa caatcgcaac ctacgagga gtaccaatt atgatcgag gcgcggcgaa 1500
taggaacttc cgcctccta ctcccagggt ctatacnaac tgggtcaaca ccatcnagac 1560
aaacggactt ttgaagctag cctcccagaa cctgttcggc atcctgtctg tggactgcac 1620
ctcccgagaa atgaattgct ttctcgagct ggtgccagga caggctggac agaaacagat 1680
cctgctcgat gccattgaca agatcgccga cgaactgggt aatcgccacc cctgcccaca 1740
cgccctctcg gtggctcccc cacaggggcc tatccctatg accgttaggt tcattagggg 1800
actgggggtg ccccgcaaac gccagatgga gccagcattt gaccaattta ggcagaccta 1860
cagacagtgg atcatcgaa ggtagcga ggggattaaa gtcaltgacg gaaagcccaa 1920
ggcacagaa atcaggcagg gggccaagga accataccct gactttgtcg acaggctctt 1980
gtccagatct aaatccgaa ggcaccctca ggagatctcc aagtcttga cagacacact 2040
gactatccaa aatgcaaatg aagagtgcag aaacgccatg aggcacctca gactgaaga 2100
taccctggag gagaaaaatg acgcatgtcg cgacattggc actaccaagc aaaagatgat 2160
gctgctcgcc aaggctctcg aaacggcctt ggtggtcca ttcaaaggag gagcaactga 2220
gggaggtcca ttgaagctg cacaacatg ttataattgt gggaaagcag gacttttato 2280
tagtcaatgt agagccctca aagtctgttt taaatgtaaa cagcctggac attttccaaa 2340
gcaatgcaga agtgttccaa aaaacgggaa gcaaggggct caagggaggg cccagaaaaa 2400
aactttcccg atacaacaga agagtgcaga caacaaatct gttgtacaag agactctcca 2460
gactcaaaat ctgtaccctg atctgagcga aataaaaaag gaatacaatg tcaaggagaa 2520
ggatcaagta gaggatctca accctggacag tttgtgggag taacatacaa tctcgagaag 2580
aggccacta ccatctgctc galcaatgac accctcttta atgtgtgctt ggcacacgga 2640
gccgacacca gcgttctcac tactgtctac tataacagac tgaatacag aggaaggaaa 2700
taccagggca caggcatcat cggcgttggg ggcacagctg aaaccttttc cactctgtc 2760
accatcaaaa agaaggggag acacattaaa accagaatgc tgggtgcgga catcccgctc 2820
accatccttg gcagagacat tctccaggac ctgggcgcta aactcgtgct ggcacaaactg 2880
tctaaggaaa tcaagtctcg caagatcgag ctgaaagagg gcacaatggg tccaaaaatc 2940
ccccagtggc cctcgaccaa agagaagctt gagggcgcta aggaaatctt gcagcgcctg 3000
ctttctgagg gcaagattag cgaagccagc gacaataacc cttacaacag ccccatcttt 3060
gtgattaaag aaaggagcgg caaatggaga ctctgcagg accctgaggga actcaacaa 3120
accgtccagg tcggaactga gatctctcgc ggaactgcctc accccggcgg cctgattaaa 3180
tgcaagcaca tgacagtcct ttttaccatc cccagctcat ttaccatccc cctcgatcct 3240
gaatttcgcc cctatactgc ttttaccatc cccagctcat atcaccagga gcccgataaa 3300
cgctatgtgt ggaagtgcct cccccaggga tttgtgctta gcccctacat ttaccagaa 3360
acacttcaag agatcctcca acccttccgc gaaagatacc cagaggttca actctaccaa 3420
tatatggacg acctgttcat ggggtccaac ggggtctaaga agcagcaca ggaactcatc 3480
atogaactga gggcaatcct cctggagaaa ggcttcgaga caccgcagca caagctgcaa 3540
gaagttcttc catatagctg gctgggctac cagctttgcc ctgaaaaactg gaaagtccag 3600
aagatgcagt tggatatggt caagaaccca acactgaacg acgtccagaa gctcatgggc 3660
aatattacct ggatgagctc cggaaatccct gggcttaccg ttaagcacat tgcgcgaact 3720
acaaaaggat gctcggagtt gaaccagaag gtcatttggg cagaggaagc tcagaaggaa 3780
ctggaggaga ataataaaaa gattaagaat gctcaagggc tccaatacta caatccgaa 3840
gaagaaatgt tgtgcgaggt cgaaatcact aagaactacg aagccaccta tgtcatcaaa 3900

```

```

cagtcaccaag gcatcttctg ggcgggaaag aaaatcatga aggcccaaaa aggcctggcc 3960
accgttaaaa atctgatgct cctgtccag caggtccca ccgagtctat caccgcgtc 4020
ggcaagtgc ccaccttcaa agttcccttc actaaggagc aggtgatgtg ggagatgcaa 4080
aaaggctggt actactcttg gttcccgag atcgcttaca cccaccaagt ggtgcacgac 4140
gactggagaa tgaagcttgt cgaggagccc actagcggaa ttacaatcta taccgacggc 4200
ggaaagcaaa acggagaggg aatcgctgca tacgtcacat ctaacggccg caccaagcaa 4260
aagaggtctg gcctgtcac tcaccaggtg gctgagagga tggctatcca gatggccctt 4320
gaggacacta gagacaagca ggtgaacatl gtgactgaca gctactactg ctggaaaaac 4380
atcacagagg gccttggcct ggagggaccc cagtctccct ggtggcctat catccagaat 4440
atccgcgaaa aggaaattgt ctatttcgcc tgggtgcctg gacacaaagg aatttacggc 4500
aaccaactcg ccgatgaagc cgccaaaatt aaaggagaaa tcatgcttgc ctaccagggc 4560
acacagatta aggagaagag agacgaggac gctggcttbt accgtgtgtg gccatacgac 4620
atcatgattc ccgttagcga cacaagatc attccaaacc atgtcaagat ccaggtgcca 4680
cccaattcat ttggttgggt gacgggaaag tccagcatgg ctaagcaggg cttctctgatt 4740
aacgggggaa tcattgatga aggatcaccc ggcgaaatcc aggtgatctg cacaaatata 4800
ggcaaaagca atattaagct tatcgaaagg cagaagtctg ctcaactcat catcctccag 4860
cacacacgca attcaagaca accctgggac gaaaacaaga ttacccagag aggtgacaaq 4920
ggcttcggca gcacaggtgt gttctgggtg gagaacatcc aggaagcaca ggacgagcac 4980
ggaattctgc acactccccc taagattttt qcccgcaatt ccaagatccc actgactgtg 5040
gctaagcaga tcacacagga atgccccac tgcacaaaac aaggttcttg cccgcgcggc 5100
tgctgtatga ggtcccccag tcactggcag gcagattgca cccactcga caacaaaatt 5160
atcctgacct tctgtggagag caattccggc tacatccag caacactcct ctccaaaggaa 5220
aatgcatgtt gcacctccct cgaattctg gaattgggca ggctgtcttc tccaaaatcc 5280
ctgcaacacg acacaggcac caactttgtg gctgaaactg tgggtgaatc gctgaagtcc 5340
ctgaaatctg ccacacacac tggcattccc tatcccttg aaagccaggg catctgtcag 5400
agggccaaac gaactctgaa agaaaagatc caatctcaca gagacaatac acgacattg 5460
gagggccgac ttcatctgc ccttatcacc tgcacaaaag gaagagaaag catggcgagg 5520
cagaccctct ggaggtctt catcactaac caggccaggg tcatccatga aaagctgctc 5580
ttgcagcagg ccagatcttc caaaaagttc tgcctttata agatcccggt tgagcacgac 5640
tggaaagtgc acactccccc ttgtggaaa ggagacggcg cagtgtggtt gacgatgag 5700
ggcaagggga tcactgctgt gcccttgaca cgcaccaaagc ttctcatcaa gccaaactga 5760
accggggggg gcgctctccc tttagtgagg gttaatgctt cagcagaca tgataagata 5820
cattgatgag ttggacaaa aatgcagtga aatgcagtga aaaaaatgct ttattttgtg 5880
aatttgtgat gctattgctt tatttgtaac cattataagc tgcaataaac aagttaacaa 5940
caacaattgc attcatttta tgtttcagggt tcagggggag atgtgggagg ttttttaag 6000
caagtaaaac ctctacaaat gtggtaaaat ccgataagga tccatccggg ctggcgtaat 6060
agcgaagagg ccgcacccga tgcaccttc caacagttgc gcagcclgaa lggcgaatgg 6120
acgcgcctcg tagcgcgca ttaagcgcg cggtgtgtgt ggttaacgag agcgtgacgg 6180
ctacacttgc cagcgcccta gcgcgcctc ctttgccttt cttcccttcc ttctcgcca 6240
cgttcgcagg ctttccctcg caagctctaa atcgggggct ccttttaggg ttccgattta 6300
gagcttlaag gcacctcgac cgcacaaaac ttgatttggg tgatggttca cgtagtgggc 6360
catcgccctg atagacggtt ttccgcctt tgacgttggg gtccacgttc tttaatagt 6420
gactcttgtt ccaactgga acaacactca accctatttc ggtctattct tttagttat 6480
aagggttttt gcgatttcg gctatttggg taaaaaatga gctgatttaa caaatattta 6540
acngaatatt taacaaaata ttaacgttta caatttcgcc tgaatgggta tttctcctt 6600
acgcactctg gcggtatttc acaccgata cgcggtctg cgcagacca tggcctgaaa 6660
taacctctga aagaggaaat tggttaggta ccttctgagg cggaaagaac cagctgtgga 6720
atgtgtgtca gttagggtgt gaaaagtccc caggctcccc agcagggcaga agtatgcaaa 6780
gcattgcatc caattagtca gcaaccaggt gtgaaaagtc cccaggtccc ccagcaggca 6840
gaagtatgca aagcatgcat ctcaattagt cagcaaccat agtcccgccc ctaactcggc 6900
ccatcccgcc cctaaactcg cccagttccg cccattctcc gcccatggc tgactaattt 6960
ttttatttta tgcagaggcc gagggcgccct cggcctctga gctattccag aagttagtag 7020
gaggtttttt tggaggccca ggtttttgca aaaagcttga ttcttctgac acaacagctc 7080
cgaacttaag gctagagcca ccatgattga acaagatgga ttgcaacgag gttctcggc 7140
cgcttgggtg gagaggetat teggtatga ctgggcaaaa cagacaatcg gctgctctga 7200
tgccgcgtg ttccggctgt cagcgacagg gcgcccgtt ctttttgtca agacgcacct 7260
gtccggtgcc ctgaatgaac tgcaggcga ggcagcgagg ctatcgtagg tggccacgac 7320
ggcgcltccc tgcgacgtg tgcctgacgt tgcactgaa gcgggaaagg actggclgct 7380
attggcgaaa gtgcggggc aggatctcct gtcactcac cttgctcctg ccagaaaagt 7440
atccatcatg gctgatgcaa tgcggcggt ccatcgcctt gatccgcta cctgoccat 7500
cgaccacaaa gcgaacatc gcatcgagcg agcacgtact cggatggaag ccggtcttgt 7560
cgatcaggat gatctggac aagagcatca ggggtctcgg ccagccgaac tgttcggcag 7620
gctcaaggcg ccatgcctg acggcgagga tctcgtctgt acccatggcg atgctgctt 7680
gccgaatata atgggtgaaa atggccgctt ttctggattc atcgactgtg gcggctggg 7740

```

```

tgtggcggac cgctatcagg acatagcgtt ggctaccggt gatattgctg aagagcttgg 7800
cggcgaatgg gctgacgcgt tctcgtgct ttacgggtat gccgctcccg attcgacgg 7860
catcgctctc tatcgctctc ttgacgagtt cttctgagcg ggactctggg gttcgaaatg 7920
accgaaccaag cgacgcceaa cctgccatca cgtatggccg aataaaatat ctttattttc 7980
attacatctg tgtgttggtt ttttgtgtga atcgatagcg ataaagatcc gcgtatggtg 8040
caactctcagt acaatctgct ctgatgccgc atagttaagc cagccccgac accgcceaac 8100
accogctgac ggcgcctgac gggcttgtct gctcccgcca tccgcttaca gacaagctgt 8160
gacgcgtctcc gggagctgca tgtgtcagag gttttccacc tcatcaccga aacgcgcgag 8220
acgaagggc ctcgtgatac gcctattttt ataggttaat gtcatgataa taatgggttc 8280
ttagacgtca ggtggcactt ttccgggaaa tgtgcgcgga acccctattt gtttattttt 8340
ctaaatacat tcaaatatgt atccgctcat gagacaataa cccgtataaa tgcctcaata 8400
atattgaaaa aggaagagta tgagtattca acatttccgt gtgcctctta ttcccttttt 8460
tgccggcattt tgccttctct tttttgctca cccagaaaac ctgggtgaaag taaaagatgc 8520
tgaagatcag ttgggtgcac gagtgggtta catcgaaact gatctcaaca gcggttaagat 8580
ccttgagagt tttcgccccg aagaacgttt tccaatgat agcactttta aagttctgct 8640
atgtggcgcg gtattatccc gtattgacgc cgggcaagag caactcggtc gccgcataca 8700
ctattctcag aatgacttgg ttgagtactc accagtcaaa gaaaagcatc ttacggatgg 8760
catgacagta agagaattat gcagtgtctc cataaccatg agtgataaca ctgcggccaa 8820
cttacttctg acaacgatcg gaggaccgaa ggagctaacg gcttttttgc acaacatggg 8880
ggatcatgta actcgcttg atcgttggga accggagctg aatgaagcca taccaaacga 8940
cgagcgtgac accacgatgc ctgtagcaat ggcaacaaac ttgcgcaaac tattaactgg 9000
cgaactactt actctagctt cccggcaaca attaatagac tggatggagg cggataaagt 9060
tgcaggacca cttctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg 9120
agccggtgag cgtgggtctc gccgtatcat tgcagcactg gggccagatg gtaagccctc 9180
ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca 9240
gatcgctgag ataggtgect cactgattaa gcattggtaa ctgtcagacc aagtttactc 9300
atatatactt tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat 9360
cctttttgat aatctcatga ccaaaatccc ttaacgtgag ttttctgtcc actgagcgtc 9420
agaccccgta gaaaagatca aaggatcttc ttgagatcct tttttctgc cgtaatctgc 9480
ctgcttgcaa acaaaaaaac caccgctacc agcgttgggt tgttttgcgg atcaagagct 9540
accaactctt tttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct 9600
tctagtgtag cctagtgtag gccacnactt caagaactct gtagcacgcg ctacataact 9660
cgctctgcta atcctgttac cagtggctgc tgcagtggc gataagtcgt gtccttacgg 9720
gttggaacta agacgatagt taccggataa ggccgcagcg tccgggtgaa cgggggggtc 9780
gtgcacacag cccagcttgg agcgaaacgac ctacaccgaa ctgagatacc tacagcgtga 9840
gctatgagaa agcgcacgcg ttcccgaagg gagaagcgcg gacaggtatc cggtaagcgg 9900
cagggtcgga acaggagagc gcacgagggg gcttccaggg ggaacgcctt ggtatcttta 9960
tagtccctgtc gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg 10020
ggggcggagc cttatgaaaa acgccagcaa cgcggccttt ttacggttcc tggccttttg 10080
ctggcctttt gctcacatgg ctcgacagat ct 10112

```

<210> 6

<211> 10227

<212> DNA

<213> Artificial Sequence

<220>

<223> LpESYNGP

<400> 6

```

tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattgggta 60
ttggccattg catacgittgl alclatatca taatatgtac alttatattg gctcatgtcc 120
aatatgacgg ccattgttggc attgattatt gactagttaa taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaaa taacttaacg taaatggccc 240
gcctggctga ccgcccaacg acccccgccc attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtattttac ggttaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagtcg cccctatttg acgtcaatga 420
cggtaaatgg cccgcctggc attatgccca gtacatgacc ttacgggact ttccactgtg 480
gcagtaacac tacgtattag tcatcgtat taccatgggt atgcggtttt ggcagtacac 540
caatgggcgt gगतacgggt ttgactcaag gggatttcca agtctccacc ccattgacgt 600
caatgggagt ttgttttggc accaaaaatc acgggacttt ccaaaatgtc gtaacaaactg 660
cgatcgcccg cccgcttgac gcaaatgggc ggtaggcgtg tacggtggga ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780

```

```

agttaaattg ctaacgcagt cagtgcctct gacacacag tctcgaactt aagctgcagt 840
gactctctta aggtagcctt gcagaagtgg gtcgtgaggg actgggcagg taagtatcaa 900
ggttacaaga caggtttaag gagaccaata gaaactgggg ttgtcgagac agagaagact 960
cttgcgtttc tgataggcac ctattgggtc tactgacatc cactttgcct ttctctccac 1020
aggtgtccac tccaggttca attacagctc ttaaggctag agtacttaat acgactccact 1080
ataggctaga gaatttcaga ggggcgcaga ccctacctgt tgaacctggc tgatcgtagg 1140
atccccggga cagcagagga gaacttacag aagtcttctg gaggtgttcc tggccagaac 1200
acaggaggac agttaagatg ggcgatcccc tcacctggtc caagccctg aagaaactgg 1260
aaaaagtcac cgttcagggg agccaaaagc ttaccacagg caattgcaac tgggcatctgt 1320
ccctggtgga tcttttccac gacactaatt tcgttaaggga gaaagattgg caactcagag 1380
acgtgatccc cctcttgagg gacgtgaccc aaacattgtc tgggcaggag cgcgaagctt 1440
tcgagcgcac ctgggtgggc atcagcgccg tcaaaatggg gctgcgaatc aacacagctg 1500
ttgacggtaa agctagcttt caactgctcc gcgctaagta cgagaagaaa accgccaaca 1560
agaaacaatc cgaacctagc gaggagtacc caattatgat cgacggcgcc ggcaatagga 1620
acttcgcgcc actgactccc aggggctata ccacctgggt caacaccatc cagacaaacg 1680
gacttttgaa cgaagcctcc cagaacctgt tcggcatcct gtctgtggac tgcacctccg 1740
aagaaatgaa tgcttttctc gacgtgggtg caggcacagg tggacagaaa caqatcctgc 1800
tcgatgccat tgacaagatc gcgcagcact gggataatcg ccacccctcg ccaaacgccc 1860
ctctggtggc tccccacag ggcctatccc ctatgacgc taggttcat aggggactgg 1920
gggtgcctcg cgaacgcag atggagccag catttgacca atttaggcag acctacagac 1980
actggatcat cgaagccatg agcgaaggga gatcgaaaag cccaaggcac 2040
agaacatcag gcaaggggcc aaggaacct accctgagtt tgcgacagg ctctgtctcc 2100
agattaaatc cgaaggccac cctcaggaga tctccaagtt ctgacagac acactgacta 2160
tccaaatgc aatatgagg tgcaaaaacg ccatgaggca cctcagacct gaagataccc 2220
tggaagcaaa aatgtacgca tgcgcacaca ttggcactac caagcaaaag atgatgctgc 2280
tcgccaagcc tctgcaaaac ggctggctg gtccattcaa aggaggagca ctgaagggag 2340
gtccattgaa agctgcacaa acatgttata attgtgggaa gccaggacat tbatctagtc 2400
aatgtagagc acttaaatgc tgttttaaat gtaaacagcc tggacatttc tcaaaccaat 2460
gcgaagtgat tccaaaaaac ggaagcaag gggctcaagg gaggccccag aaaaaactt 2520
tccgatatac acagaaggat cagcacacaa aatctgttgt acaagagact cctcagact 2580
aaaaatctga ccagatctg agcgaataa aaaaagaaat caatgtcaag gagaaggatc 2640
aagtacagga tctcaacctg gacagtttgt gggagtaaca tacaatctcg agaagaggcc 2700
cactaccac ctcctgatca atgacacccc tcttaatgtg ctgctggaca cggagccga 2760
cacacggct ctcactactg ctactataaa cagactgaaa tacagaggaa ggaataacca 2820
gggcacagcc atcatcgcg ttggaggcaa cgtcgaaacc ttttccactc ctgtcaccat 2880
caaaaagaa gggagacaca ttaaacccag aatgctgtgc gcgacatcc ccgtcaccat 2940
ccttgcgaga gacattctcc aggacctlgy cyctaaactc gtgctlgycc aactgtctaa 3000
ggaaatcaag ttccgcaaga tcgagctgaa agagggcaca atgggtccaa aaatccccca 3060
gtggccccc accaaagaga agcttgaggg cgttaaggaa atcgtgcagc gcttgccttc 3120
tgaggccaag attagcgagg ccagcgacaa taaccttac aacagcccca tctttgtgat 3180
taagaaaagg agcggcaaat ggagactcct gcaggacctg agggaaactc acaagacngt 3240
ccaggtcgga actgagatct ctgcgcgact gcctcacccc ggcggcctga ttaaatgcaa 3300
gcacatgaca gtccctgaca ttggagacgc ttattttacc atccccctg atcctgaatt 3360
tcgcccctat actgctttta ccatccccag calcaatcac caggagcccg ataaacgcta 3420
tgtgtggaag tgcctcccc agggat.ttgt gcttagcccc tacatttacc agaagacact 3480
tcaagagatc ctccaacctt tccgcgaaag ataccagag gttcaactct accaatatat 3540
ggacgcctcg ttcattgggt ccaacgggtc taagaagcag cacaaggaa ctcactatga 3600
actgagggca atcctcctgg agaaaggctt cgagacaccc gacgacaagc tgcagaagat 3660
tccctccat agctggctg gctacacagct ttgcccga aactggaag tccagaagat 3720
cgagttggat atggctcaaga acccaacact gaacgaagct cagaagctca tgggcaatat 3780
tacctggatg agctccgga tccctgggtc taccgttaag cacatgcgc caactacaaa 3840
aggatgcgc gagttgaacc agaaggtcat ttggacagag gaagctcaga aggaactgga 3900
ggagaataat gaaaagatta agaattgcta agggctccaa tactacaa ccaagaga 3960
aatgtgtgct gaggtcgaaa tcaactaagaa ctacgaagcc acctatgtca tcaaacagtc 4020
ccaaggcac ttgtgggcg gaaagaaaat catgaaggcc acaaaaggot ggtccaccgt 4080
taaaaatctg atgctcctgc tccagcacgt cgccaccgag tctatcccc cgtcggcga 4140
gtgccccacc ttcaaatgtc ccttcaacta ggaagcaggt atgtgggaga tgcaaaaagg 4200
ctgtactlac tcttggcllc ccgagatcgt clacacccac caagtgtgtc acgacgactg 4260
gagaatgaag cttgtcgagg agccactag cggattaca atctataccg acggcgga 4320
gcaaaacgga gagggaaatc ctgcatacgt cacatctaac ggccgcacca agcaaaagag 4380
gctcgccct gtcaatcac aggtggctga gaggtggct atccagatg cccttgagg 4440
cactagagca aagcaggtga acatttgtgac tgacagctac tactgctgga aaaaatcac 4500
agagggcctt ggcctggag gacccagtc tccctgggtg cctatcatcc agaatatcc 4560
cgaaaaggaa attgtctatt tcgctgggtt gcctgggac aaaggaattt acggcaacca 4620

```

actgcccgat gaagccgcga aaattaaaga ggaatcatg cttgcctacc agggcacaca 4680
gattaaggag aagagagacg aggcagctgg cttgacctg tgtgtgcat acgacatcat 4740
gattccccgt aggcacacaa agatcattcc aaccgatgtc aagatccagg tgcaccccaa 4800
ttcatttggg tgggtgacgc gaaagtccag catggctaag cagggtcttc tgattaacgg 4860
gggaattcatt gatgaaggat acaccgcgga aatccagggt atctgcacaa atactggcga 4920
aagcaatatt aagcctatcg aagggcagaa gtctgctcaa ctcatcatcc tccagcacca 4980
cagcaattca agcaacactt gggacgaaaa caagattagc cagagaggtg acaagggtt 5040
cggcagcaca ggtgtgtctt ggttgagaaa catccaggaa gcacagagaa 5100
ttggcaccac tcccttaaga ttttggcccg caattacaag atcccactga ctgtgtgctaa 5160
gcagatcaca caggatgcc cccactgcac caacaagggt totggcccg ccggtcgtgt 5220
gatgaggtcc cccatcact ggcaggcaga ttgcacccac ctgcacaaac ctccctcca 5280
gaccttctgt gagagcaatt ccggtacat ccacgcaaca ctccctcca aggaaaaatgc 5340
attgtgcacc tccctcgcaa tctggaaatg ggcaggctg tctctccaa aatccctgca 5400
cacgcacaac ggcaccaact ttgtggctga acctgtgggt aatctgctga agttccctgaa 5460
aatgcgccac accactggca ttccctatca cctgaaaagc cagggtattg tgcagagggc 5520
caacagaact ctgaagaaaa agatccaatc tcacagagac aatacacaga cattggaggc 5580
cgcaactcag ctgcctctta tcacctgcaa caaaggaaga gaaagcatgg gcggccagac 5640
ccccctggag gtcttcatca ctaaccaggc ccaggctcat catgaaaaag tgcctctgca 5700
gcagggccag tccctcaaaa agttctgtct ttataagatc cccggtgagc acgactggaa 5760
aggtccctaca agagtttctt ggaaggaga cggcgcatgt gtggtgaacg atgagggcaa 5820
ggggtatcgc gcgtgtccgc ljacacgcac caagctctc alccagccaa actgaacccg 5880
gggcggccgc ttcctcttag tgggggttaa tgcctcgagc agacatgata agatacattg 5940
atgagtttgg acaaacacaa actagaatgc agtgaaaaaa atgettattt tgtgaattt 6000
gtgatcctat tgccttattt gtaaccatla taagctgcaa taaacaagt taaacaacaa 6060
attgcatcra ttttatgttt cagggttcagg gggagatgtg gggggtttt taaagcaagt 6120
aaaaactcra caaatccgat aaaatccgat aaggatcgat ccgggctggc gtaatagcga 6180
agagggccgc accgatgcc ctcccaaca gttgcgcagc ctgaatggcg aatggagcgc 6240
ccctgttagc gctcattagc cgcggcgggt gtggtgtgta cgcgcagcgt gacccgctaca 6300
cttgccagcg ccttagcgcc cgtctcttcc gcttcttccc ctctcttct cgcacagctc 6360
gcgggcttcc cctgcaagc tctaaatcgg ggggtccctt tagggttccg atttagagct 6420
ttacggccac tgcacggcaa aaaaactgat ttgggtgatg gtctacgtag tgggccaatg 6480
ccctgataga cgggttttcc ccttttgacg ttggagtcca cgttctttaa tagtggactc 6540
ttgttccaaa ctggaacaa actcaacct atctcgttct actcttttga ttataaggg 6600
atttgcgaa tttcggccta ttggttaaaa aatgagctga tttaacaaat atttaacgg 6660
aattttaaca aatatatac gtttacaatt tccgtgatg cgggtatttc tccctacgca 6720
tctgtccggt atttccaccc gcatacggcg atctgcgcag caccatggcc tgaataaacc 6780
tctgaagag gaactgtgtt aggtaccttc tgaggcggaa agaaccagct gtggaatgtg 6840
tgtcagttag ggtgtggaaa gtccccaggc tcccagcag gcagaagtat gcaaaagcatg 6900
catctcaatt agtcagcaac cagggtgtga aagtcccgag gctcccagc aggcagaagt 6960
atgcaagaca tgcattcaca ttagtcagca accatagtcc cgcctcaac tccgccaatc 7020
cgcctcctaa ctcgcccag cctcggccat tctcggccat atggctgact aatttttttt 7080
atttatgcag aggcggagcg cgcctcgccc tctgagctat tccagaagta gtgagaggc 7140
tttttggag gctatggctt ttgcaaaaag cttgattctt ctgacacaa agtctcgaa 7200
ttaaggctag agccaccatg attgaacaa atggattgca cgcaggttct cccgcccgtt 7260
gggtggagag gctattcggc tatgactggg cacaacagac aatcggctgc tctgatgcg 7320
ccgtgttccg gctgtcagcg caggggcgcc cggttctttt tgtcaagacc gaactgtccg 7380
gtgcctcga tgaactcgag gacgagcgag cgcggctatc gtggctggcc acgacggcg 7440
ttccttgcgc agctgtgttc gacgtgtcga ctgaagcggg aagggaatgg ctgctattg 7500
gcgaagtgcg ggggcaggat ctctgtcat ctcacctgc tccgtcgag aaagtatca 7560
tcactggctga tgaattgcgg cggctgcata cgcctgcatc gcttcatcc ccatcgacc 7620
accaagcgaa acatcgcatc gaggcgagcac gtaactggat ggaagccggt ctgttcgatc 7680
aggatgatct ggcagcaagag catcaggggc tccgcccagc cgaactgttc gccaggctca 7740
agggcgcat gcccgagcg gaggatctcg tctgaccca tggcgatgcc tgcctgcga 7800
atatcctggg ggaatatggc cgttttctg gattcatcga ctgtggccg ctgggtgtg 7860
cggacggcta tcaagacata cgttggcta cccgtgatc tgcgaagag cttggccg 7920
aatgggctga ccgcttctc gtgctttacg gtatcgccg tcccgattcc cagcgcatg 7980
ccttctatcg ccttcttgac gaggttctct gagcgggact ctggggttcg aaatgacga 8040
ccaagcgagc ccaacctgc catcacgatg gccgcaataa aatatcttta ttttattac 8100
atctgtgtgt tggttttttg tgtgaatcga tagcgataag gatcccgta tgggtgcatc 8160
tcagtacaat ctgctctgat gccgatagt taagccagcc cgcacaccg ccaacaccg 8220
ctgacgcgc ctgacgggct tgtctgtccc cggcatcgc ttacagacaa gctgtgacg 8280
tctccgggag ctgcatgtgt cagaggtttt caccgtcatc accgaaaac gcgagacga 8340
agggcctcgt gatccgcta ttttatagg ttaatgtcat gataataatg gtttcttga 8400
cgtcaggctg cacttttcg ggaatgtgc gcggaacccc tatttgttta ttttctaaa 8460

tacattcaaa tatgtatccg ctcatgagac aataaccccg ataatgtct caataatatt 8520
gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc ccttattccc ttttttgcg 8580
cattttgctc tctgtttttt gctcacccag aaacgctggt gaaagtaaaa gatgtgaag 8640
atcagtttgg tgcaagagtg ggttacatcg aactggatct caacagcggg aagatccttg 8700
agagttttcg ccccgaaaga cgttttccaa tcatgagcac ttttaaaat ctgctatgtg 8760
gcggcgatatt atccgtattt gacgcggggc aagagcaact cggctcgcgc atacactatt 8820
ctcaagatga cttggttagg tactcacagc tcacagaaaa gcatcttacc gatggcatga 8880
cagtaagaga attatcgagt gctgcataa ccatgagtga taacactgcg gcaacttac 8940
ttctgacaa gatcggagga ccgaaggagc taaccgtttt ttgtcacaa atgggggac 9000
atgtaactcg ccttgatcgt tgggaacccg agctgaatga agccatacca aacgacgagc 9060
gtgacccac gatgcctgta gcaatggcaa caacgttgcg caaactatta actggcgaa 9120
tacttactct agcttcccg gcaacaattaa tagactggat gggagcggat aaagtgtcag 9180
gaccacttct gcgctcgccc ctccggctg gctggtttat tgcgtataaa tctggagccg 9240
gtgagcgttg gtctcgcggt atcattgcag cactggggcc agatgtgtaa cctcccgta 9300
tctagttat tccacagagc gggagtcag caactatgga tgaacgaaal agacagatcg 9360
ctgagatagg tgcctcactg attaagcatt ggttaactgt agaccaagt tactcatata 9420
tactttagat tgatttaaaa cttcattttt aatttaaaag gatctagggt aagatcctt 9480
ttgataatct catgacaaaa atcccttaac gtgagtttcc gtccactga gcgtcagacc 9540
ccgtagaaaa gatcaaggga tcttcttgag atcctttttt tctgcgcgta atctgctgct 9600
tgcaaacaaa aaaaccaccc ctaccagcgg tggtttgttt gccggatcaa gactaccaa 9660
ctctttttcc gaaggttaact ggtctcagca gagcgagat accaaatact gtcctctag 9720
tgtagccgta tttagccac caottcaaga actctgtagc accgactaca taccctgctc 9780
tgctaatcct gttaccagt gctgctgcca gtggcgataa tgcgtgtctt accgggttgg 9840
actcaagagc atagttaccg gataaggcgc agcggctcgg ctgaacgggg ggttctgta 9900
cacagcccg cttggagcga acgacctaca ccgaactgag atacctacg cgtgagctat 9960
gagaaagcgc cagcgttccc gaagggagaa aggcggagag gtatccgcta agcggcagg 10020
tcggaacagg agagcgacg agggagcttc cagggggaaa cgcctgggat ctttatagt 10080
ctgtcggggt tgcacacctc tgacttgagc gtcgattttt gtgatgctc tcagggggg 10140
ggagcctatg gaaaaacgc agcaacggc cttttttacg gttcctggcc ttttctgag 10200
cttttgcata catggctcga cagatct 10227

<210> 7

<211> 10815

<212> DNA

<213> Artificial Sequence

<220>

<223> pESYNGPRRE

<400> 7

```

tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattgggta 60
ttggccattg catacgtttg .atctatatca taatatgtac atttataatg gctcatgtcc 120
aatatgaccg ccatgtttggc attgattatt gactagttat taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttaagg faaatggccc 240
gcctggtcta ccgcccaacg acccccggcc attgacgtca ataataagct atgttcccat 300
agtaacgcca atagggaact tccattgacg tcaatgggtg gagtatttac ggtaaaactg 360
ccacttggca gtacatcaag tgtatcatat gccaaagtcc cccctatag acgtcaatga 420
cggtaaatgg ccgccttggc attatgcccc gtacatgacc ttacgggact ttctactctg 480
cgagtacatc tcagtattag tcactgctat taccatgggt atgcggttct ggcagtacac 540
caatgggcgt ggaatagcgt ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatgggagt ttgttttggc accaaatca acgggacttt ccaaaatgta gtaacaactg 660
cgatcgcccg ccccgctgac gcaaatgggc ggtaggcgtg tacgggtggg ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780
agttaaatgt ctaacgcagt cagtgttctt gacacaaacg tctcgaaact aagctgcagt 840
gactctctta aggtagcctt gcagaagtgt gtcgtgaggg actgggcagg taagtatcaa 900
ggttacaaga cagggttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttc tgatagggca ctattggtct tactgacatc cactttgctt ttctctccac 1020
agggttccac tcccagttca attacagctc ttaaggctag agtacttaac acgactcact 1080
ataggctaga gaattcgcca ccatgggcga tccctccacc tgggtccaaag ccttgaagaa 1140
actggaaaaa gtacacgttc aaggtagcca aaagcttacc acaggcaatt gcaactgggc 1200
attgtccctg gtggtacttt tccacgacac taatttcgtt aaggagaaag attggcaact 1260
cagagacgtg atccccctct tggaggacgt gacccaaaca ttgtctgggc aggagcgcca 1320
agctttcgag cgcacctggt gggccatcag cgcagtcaaa atgggctgac aaatcaacaa 1380

```

```

cgtgggtgac ggtaaagcta gctttcaact gctccgcgct aagtacgaga agaaaaccgc 1440
caacaagaaa caatccgaac ctacccagga gtacccaatt atgacgacg gcgcgggcaa 1500
taggaacttc cgcctcaga ctcccagggt ctataccacc tgggtcaaca ccatccagac 1560
aaacggactt ttgaacgaag cctcccagaa cctgttccgg atcctgtctg tggaactgca 1620
ctccgaagaa atgaatgctt ttctcgacgt ggtgccagga caggctggac agaaacagat 1680
ctcgtcgtat gccattgaca agatcgccga cgaactgggt aatcgccacc ccttgcacaa 1740
cgccccctcg gtgggtcccc caccggggcc tatccctatg accgctaggt tcattagggg 1800
actgggggtg ccccgccgac gccagatgga gccagcattt gacccaatta ggcagacctt 1860
cagacagtggt atcatcgaa ccatgagcga ggggattaaa gtcatgatcg gaaagcccaa 1920
ggcacagaa acatggcagg gggccaagga accataccct gagtttgtcg acagggttct 1980
gtcccagatt aatccgaag gccaccccca ggagatctcc aagttcttga cagacacact 2040
gactatccaa aatgcaaatg aagagtgccg aaacgccatg aggcacctca gacctgaaga 2100
taccctggag gagaaaatgt acgcatgtcg cgacattggc actcaaacg aaaagatgat 2160
gctgtcgccc aaggcttctg aaaccggcct ggcgtgtcca ttcaaaaggg gagcactgaa 2220
gggaggtcca ttgaaagctg cacaacatg ttataattgt ggggaagccg gacatttalc 2280
tagtcaatgt agagcaccta aagtctgttt taaatgtaaa cagcctgyac atttctcaaa 2340
gcaatcgaga agtgttccaa aaaacgggaa gcaaggggct caagggaagg cccagaacaa 2400
aactttcccg atacaacaga agagtccgca caacaaatct gttgtacaag agactctcca 2460
gactcaaaat ctgtaccgac atctgagcga aataaaaaag gaatacaatg tcaaggagaa 2520
ggatcaagta gaggatctca acctggacag ttgtggggag taacatacaa tctcgagaag 2580
aggcccaact ccatcgtcct gatcaatgac accctcttta atgtgtctg taacatacaa 2640
gccgacacca gcgttctcac tactgtctac tataacagac tgaatacag aggaaggaaa 2700
taccagggca caggcatcat cgggttggga ggcaacgtcg aaaccttttc cactcctgtc 2760
accataaaa agaaaggagg acacattaaa accagaatgc ttggtcgcca catcccgtc 2820
accatccttg gcagagacat tctccagcac ctgggcccga aactcgtgct ggcacaaact 2880
tctaagcaaa tcaagtcccg caagatcgag ctgaaaaggg gcacaaatgg tccaaaaatc 2940
ccccagtggt ccttgacaaa agagaagcct gagggcgcta aggaatcgt gcagcgccgt 3000
ctttctgagg cgaagattag cgaggccagc gacaataacc cttaacaacg ccccatcttt 3060
gtgattaaag aaaggagcgg caaatggaga ctctcgacag acctgaggga actcaacaa 3120
accgtccagg tcggaactga gatctctcgc ggaactgcct accccggcgg cctgattaaa 3180
tgcaagcaca tgcagctcct tgacattgga gacgcttatt ttaccatccc cctcgatcct 3240
gaatttcgce cctatactgc ttttaccatc cccagcaca atcaccagga gcccgataaa 3300
cgctatgtgt ggaagtgcct cccccaggga ttgtgtgcta gccctacat ttaccagaag 3360
acacttcaag agatctccca accttccgcg gaaagatacc cagaggttca actctacaa 3420
tatatcgacg acctgttcat ggggtccaac ggggtctaaga agcagcaca ggaactcatc 3480
atcgaactga gggcaatcct cctggagaaa ggcctcgaga caccgcagca caagctgcaa 3540
gaagtctccc catatcgctg gctgggctac cagctttgcc ctgaaaaact gaaagtccag 3600
aagatgcagt tggatatggt caagaaccga acactgaacg acgtccagaa gctcatgggc 3660
aatattacct ggatgagctc oggaatccct ggggttacog ttaagcacat tgccgcaact 3720
acaaaaggat gctcggagtt gaaccagaag gtcattttga cagaggaaag tcagaaggaa 3780
ctggagcaga ataatagaaa gattaagaat gctcaagggc tccaatacta caatcccgaa 3840
gaagaatagt tgtgcgaggt cgaatcact aagaactacg aagccacctt tgtcatcaa 3900
cagtcocaa gcatctttg ggcgggaaag aaaaatcatg aggcacaa aggcctggctc 3960
accgttaaaa atctgatgct cctgctccag cacgtcgcca ccgagtctat cacccgctc 4020
ggcaagtgc ccacctcaaa agttcccttc actaaggagc aggtgatgtg ggagatgcaa 4080
aaaggtcgtg actactcttg gcttcccag atcgtctaca cccaccaag ggtgcacgac 4140
gactggagaa tgaagcttgt cgaggagccc actagcgaaa ttacaactca taccgacgca 4200
ggaaagcaaa acgggagggg aatcgtgca tactcacat ctacggccg caccaagcaa 4260
aagaggtcgc gccctgtcac tcaccaggtg gctgagagga tgggtatcca gatggccctt 4320
gaggacacta gagacaagca ggtgaacatt gtgactgaca gctaactctg ctggaaaaac 4380
atcacggagg gccctggcct ggagggaccc cagctccctt ggtggcctat catccagaat 4440
atcccgcaaa aggaatttgt ctatttcgce tgggtgcctg gacacaaag aatttacggc 4500
aaccacctcg ccgatgaag cgcacaaatt aaaggagaaa tcatgcttgc ctaccagggc 4560
acacagatta aggagaagag agacgaggac gctggctttg acctgtgtgt gccatacgac 4620
atcatgattc ccgttagcga cacaaagatc attccaaccg atgtcaagat ccaggtgcc 4680
cccaattcat ttggttggtg gaccggaaag tccagcatgg ctaagcaggg ctctctgatt 4740
aacgggggaa tcaattgtga aggatcaccc ggcgaaatcc aggtgatctg cacaaatctc 4800
ggcaaaagca atattaagct tatogaaggg caaaagtctg ctcaactcat catcctccag 4860
caccacagca attcaagca accctgggac gaaaacaaga tttagccagag aggtgacaag 4920
ggcttcgcca gcacaggtgt gttctcgggt gagaaatcc aggaagcaca ggacgagcac 4980
gagaattgca acactcccc taagattttg gcccgcaatt acaagatccc actgactgtg 5040
gctaagcaga tcacacagga atgcccccac tgcacccaaa aaggttctgg ccccgccgca 5100
tgcgtgatga ggtcccccac tcactggcag gcagattgca cccactcga caacaaatt 5160
atcctgacct tcgtggagag caattccggc tactccacg caacactcct ctccaaagg 5220

```

aatgcattgt gcaacctcct cgcattctcg gaatgggcca ggctgttctc tccaaaatcc 5280
 ctgcacaccc acaacggcac gctgaacctg tggggaatct gctgaagtcc 5340
 ctgaaaaatcg cccacacccc tggcattccc tatcaccctg aaagccaggg catgtgcag 5400
 agggccaacac gaactctgaa agaaaagatc caatctcaca gagaccaatc acagacattg 5460
 gagggccgac ttacagctcg ccttatccac tgcacaaaag gaagagaaag catgggcggc 5520
 cagacccctc gggaggtctt catcactaac caggcccagg tcatccatga aaagctgctc 5580
 ttgcagcagg ccagctcctc aaaaagtctc tgcctttata agatccccgg tgagcagcac 5640
 tggaaaggte ctacaaggst tttgtggaaa ggagacggcg cagttgtggt gaacgatgag 5700
 ggcaagggtc tcatcgctgt gcccctgaca cgcaccaagc ttctcatcaa gccaaactga 5760
 acccgagcaa tcccaggggg aatctcaacc cctattaccc aacagtcaga aaaaactaag 5820
 tgtgaggaga acacaatgtt tcaaccttat tgttataata atgacagtaa gaacagcatg 5880
 gcagaatcga aggaagcaag agaccaagaa atgaacctga aagaagaatc taaagaagaa 5940
 aaaaagaagaa atgactgggt gaaaaatagg atgtttctgt tatgcttagc caggggccctc 6000
 tggaaaggta ccagtggtgc agggctcctc ggagtcgtt accctgaagaa aaaattccat 6060
 cacaaacatg catcgcgaga agacacctgg gaccaggccc aacacaacat acacctagca 6120
 ggcgtgaccc gtggatcagg ggacaaatc tacaagcaga agtactccag gaacgactgg 6180
 aatggagatc cagaggggtg caacaggcgg ccaaagagct ggggtgaagt aatcgaggca 6240
 tttggagaga gctatatctc cgaagagacc aaaggggaga tttctcagcc tggggcggtc 6300
 atcaacgagc caaagaacgg ctctgggggg aacatcctc accaagggtc cttagactgt 6360
 gagatctcaa gcgaaggagg aaacatttat gactgttgca ttaaagccca agaaggaact 6420
 ctgcgtatcc ctgtctgtgg atttccctta tggctatttt gggggctggg gggcgctt 6480
 ccccttagtg agggttaatg cttcgagcag acatgataag atacattgat gaggttggac 6540
 aaaccacaac tagaatcgag tgaaaaaaat gctttatttg tgaattttgt gatgctattg 6600
 ctttatttgt aacactata agctgcaata aacaagttaa caacaacat tgacttcat 6660
 tttatgttca ggttcagggg gagatgtggg aggtttttta aagcaagtaa aacctctaca 6720
 aatgtggtaa aatcgcgata ggatcgatcc gggctggcgt aatagcgaag agggccgac 6780
 cgaacccctc tcccacagat tgcgcagcct gaatggcgaa tggacgcgcc ctgtagcgcc 6840
 gcattaagcg cggcggtgtg ggtggttaag cgcagcgtga ccgctacaat tgcagcgccc 6900
 ctagcgcccg ctctctctgc tttcttccct tctttctcgg ccaggttcgc cggctttccc 6960
 cgtcaagctc taaactgggg gctcctctta gggttccgat tttagacttt acggcacctc 7020
 gacgcacaaa aacttgattt gggatgaggt tcaagtagtg ggcctagccc ctgatagac 7080
 gtttttgcgc ctttgacgtt ggagctccag ttctttaaata gtggactctt gttccaaact 7140
 ggaaacaacac tcaacctat ctcggtctat tataagggat tatggcgatt tttggcgatt 7200
 tccggctatt ggttaaaaaa tgagctgatt taaacaaat ttaacgcgaa ttttaacaaa 7260
 atataaagct ttacaatttc gctgagtcgg gtattttctc cttaacgcat tbtgcggtat 7320
 ttacacacgc atacgcggat ctgcgcagca ccatggcctg aaataacctc tgaagagga 7380
 acttggttag gttactctcg agggcgaaag aacacagctg ggaatgtgtg tcaagttagg 7440
 tgttgaagat ccccgagctc cccagcaggg agaagtatgc aaagcatgca tctcaattag 7500
 tcaagcaaca ggtgtgaaaa gtccccagg tccccagcag gcagaagtat gcaagcatg 7560
 catctcaatt agtcagcaac catagtcccg cccctaacct cgcocatccc gccctaacct 7620
 ccgcccagtt ccgcccact tccgcccact ggctgaacta ttttttttat ttatgcagag 7680
 gccagggccc cctcggccct tgagctattc cagaagtagt gaggaggctt ttttggaggc 7740
 ctagggtttt gcaaaaagct tgattctctc gacacacag tctcgaaact aaggctagag 7800
 ccacctgat tgaacagat ggattgcagc cagggttccc ggcgcttgg gtggagaggc 7860
 tatctcgcta tgactgggca caacacacaa tccgctgctc tgatgcgcgc gtgttccgca 7920
 tgtcagcga cggggccccc gttctttttg tcaagacaga cctgtccggt gccctgaatg 7980
 aactgcagga cgaagcgagc cggctatcgt ggcctggccc gacggggctt ccttgccgag 8040
 ctgtgctcga cgttctcact gaagcgggaa gggactggct gctatgggg gaagtgccgg 8100
 ggcagagatc cctgtcatct caccttgctc ctgcccagaa agtatccatc atggctgatg 8160
 caatgcggcg gctgcagac cttgatccgg ctacctgccc attcgaccac caagcgaaac 8220
 atcgcatcga gcgagacagt actcgagtag aagccggctc tbtgatcag gatgatctgg 8280
 acgaagagca tcaaggcgctc gcgccagcgg aactgltcgc caggctcaag gcgcgcaltc 8340
 ccgacggcga ggaatcgctc gtgacccatg gcgatgcctg cttgcccgaat atcatggtag 8400
 aaaaatggcg cttttctgga ttcatcgact gtggccggct ggggtgtggc gaccgctatc 8460
 aggaatagc gttggctacc cgtgatattg ctgaagagct tggcgcgcaa tgggctgacc 8520
 gcttctcgt gctttacggt atcgccgctc ccgattcgca ggcgatcgcc ttctatcgcc 8580
 ttctgagca gttctctcga cgggactct ggggttcgaa atgacagacc aagcagcgc 8640
 caacctgcca tcaagatggc cgcataaaaa tatctttatt ttcattacat ctgtgtgttg 8700
 gtttttgtg tgaatcgata gggataagga tccgcgtatg gtgactctc agtaacatc 8760
 gctctgatgc cgcatagtta agccagcccc gacacccgcc aacacccgct gacgcgccct 8820
 gacgggtctg tctgtcccg goatecgctt acagacaaag tgtgacgct tccgggagct 8880
 goatgtgtca gaggttttca ccgtcatcac cgaacccgcg gagacgaaag ggcctcgtga 8940
 tacgctatt tttatagtt aatgtcatga laataatggt ttcttagaag tcaagtgcca 9000
 cttttcgggg aatgtgcgc ggaaccccta tttgtttatt tttctaaata catctaaata 9060

tgtatcgct catgagacaa taacctgat aatgcttca ataatttga aaaaggaaga 9120
 gtatgagat tcaacatttc cgtgtcgccc ttattccctt ttttgcggca ttttgcttc 9180
 ctgtttttgc tcaaccagaa acgctgggtg aagtaaaaga tctggaagat cagtgggtg 9240
 cagcagtggg ttacatcgaa ctggatctca acagcggtaa gatccttag agttttcgcc 9300
 ccgaagaacg ttttccaatg atgagcatt ttaaagtctc gctatgtggc gcggtattat 9360
 cccgtattga ccgcgggcaa gagcaactcg gtcgccgcat acactattct cagaatgact 9420
 tggttgagta ctcaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat 9480
 tatgcagtg tgccataacc atgagtgata acactgcggc caacttactt ctgacacga 9540
 tcggaggacc gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc 9600
 ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga 9660
 tgctgtagc aatggcaaca acgttgcgca aactattaac tggcgaaacta ctactctag 9720
 cttcccgcca acaattaata gactggatgg agggcgataa agttgcagga ccacttctgc 9780
 gctcggccct tccggctggc tgggttattg ctgataaatc tggagccggt gagcgtgggt 9840
 ctgcgggtat cattgacgca ctggggccag atggtaagcc ctcccgctatc gtagtatatc 9900
 acacgacggg gactcaggca actatggatg aacgaaatag acagatcgct gagataggtg 9960
 cctcactgat taagcattgg taactgtcag accaagttta ctcatatata ctttagattg 10020
 atttaaaact tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca 10080
 tgaccaaaat cctttaagct gagttttctg tccactgagc gtcagacccc gtgaaaaaga 10140
 tcaaaagatc ttcttgagat ccttttttct tgcgcgtaat ctgctgcttg caaacaacaa 10200
 aaccaccgct accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga 10260
 aggttaactgg cttcagcaga gcgcagatac caaataactg ccttctagt tagccgtagt 10320
 taggccaaca cttcaagaa cctgtagcac cgtctacata cctcgctctg ctaactcgtg 10380
 taccagtgcc tgcgtccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat 10440
 agttacogga taaggcgagc cggctgggct gaacgggggg tctgtgcaca cagcccagct 10500
 tggagcgaac gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca 10560
 cgcttccgga agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacagga 10620
 agcgacagag ggaacttcca gggggaaacg cctggtatct ttatagtcct gtcgggttct 10680
 gccacctctg acttgagcgt cgatttttgt gatgtctctc agggggggcg agcctatgga 10740
 aaaaacgag caacggggc tttttacggt tcttggcctt ttgctggcct tttgtcaca 10800
 tggctcgaca gatct 10815

<210> 8

<211> 10930

<212> DNA

<213> Artificial Sequence

<220>

<223> LpESYNGPRRE

<400> 8

```

tcaatattgg ccattagcca tattattcat tggttatata gcataaacca atattggcta 60
ttggccattg catakgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
aatatgaccg ccatgttggc attgattatt gactagttaa taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttaagg faaattggcc 240
gcctggctga ccgcccaacg acccccgccc attgacgtca ataataagct atgttcccat 300
agtaacgcca atagggaact tccattgacg tcaatgggtg gagtatttac ggtaaaactg 360
ccacttggca gtacatcaag tgtatcataa gccagtcog cccctattg acgtcaatga 420
cggtaaatgg ccgcctggcg attatgccc a gtacatgacc ttacgggact tccctacttg 480
gcagtgacatc tctgtattag tcatcgctat taccatgggt atgcggtttt ggcagtacac 540
caatggcgct ggaatagcgt ttgactcacc gggatttcca agtctccacc ccattgacgt 600
caatgggagt ttgttttggc accaaatca accggacttt ccaaaatgta gtaaacactg 660
cgatcgcccg ccccgctgac gcaaatggcg ggtaggcggt tacggtggga ggtctatata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780
agttaaatgt ctaacgcagt cagtgtctct gacacaacag tctcgaaact aagctgcagt 840
gactctctta aggtagcctt gcagaagtgt gtctgtaggc actgggcagg taagtataca 900
ggttacaaga caggtttaag gagaccaata gaaactggcg ttgtcgagac agagaagact 960
cttgcgtttc tgaatggcac ctattggtct tactgacatc cactttgctt tctctccac 1020
aggctccac tcccagttca attacagctc tttaggttag agtacttaac acgactcact 1080
ataggctaga gaattcgaga gggcgccaga ccttacctgt tgaacctggc tgatcgtagg 1140
atcccgagg aagcagagga gaactacag aagtctctcg gagtggttcc tggccagaac 1200
acaggaggac aggttaagatg ggcgatcccc tcacctgggt caaacccctg aagaaactgg 1260
aaaaagtcac cgttcagggt agccaaaagc ttaccacagg caattgcac tgggcattgt 1320
ccctggtgga tcttttccac gacactaatt tcgttaagga gaaagattgg caactcagag 1380

```

```

acgtgatccc cctcttggag gacgtgaccc aaacattgtc tgggcaggag cgcgaagctt 1440
tcagagccac tgggtgggac atcagccgag tcaaatggg gctgcacac aacaactggt 1500
ttgacggtaa agctagcttt caactgctcc gcgtlaagta cgagaagaaa accccaaca 1560
agaaacaact cgaacctagc gaggagtaac caattatgat cgaaggcgcc ggcaatagga 1620
acttccgccc ctagactccc aggggctata ccactgggt caacacatc cagacaaacg 1680
gacttttgaa cgaagcctcc cagaacctgt tcggcatcct gtctgtggac tgcacctccg 1740
aagaaatgaa tgcctttctc gacgtgggtg caggacaggg tggacagaaa cagatcctgc 1800
tcgatgccat tgacaagatc gccagcagct gggataatcg ccccccctg ccaaacgccc 1860
ctctggtgtg tccccacag ggcctatcc ctatgacgc taggttctat aggggactgg 1920
gggtgccccg cgaacgcag atggagccag catttgacca atttaggcag acctacagac 1980
agtggatcat cgaagccatg agcaggggga tttaagtcat gatcggaaag cccaaggcac 2040
agaaacatcag gcaggggccc aaggaacctat accctgagtt tctcgacagg cttctgtccc 2100
agattaaatc cgaaggccac cctcaggaga tctccaagtt cttgacagac acactgacta 2160
tccaaaatgc aaatgaagag tgcagaaacg ccatgaggca cctcagacct gaagatcccc 2220
tggaggagaa aatgtatcga tctcgcgaca ttggcactac caagcaaaag atgatgtgc 2280
tcgccaaagc tctgcaaaccc ggctgggtg gtccattcaa aggaggagca ctgaaggag 2340
gtccattgaa agctgcacaa acatgttata attgtgggaa gccaggacat ttatctagtc 2400
aatgtagagc acctaaagtc tgttttaaat gtaaacagcc tggacatttc taaaagcaat 2460
gcagaagtgt tccaaaacac gggaaagcaag gggctcaagg gaggccccag aaacaaactt 2520
tccgatatac acgaagagat cagcaacaaca aatctgttgt acaagagact cctcagaact 2580
aaaatctgtc ccagatctcg agcgaataaa aaaaggaata caatgtcaag gagaaggalc 2640
aagtacagga tctcaactcg gacagtttgt gggagtaaca tacaatctcg agaagaggcc 2700
cactaccatc gtctgtgatc atgacaacccc tcttaatgtg ctgctggaca ccggagccga 2760
caccagcgtt ctaactactg ctcaactata cagactgaaa tacagaggaa ggaataacca 2820
gggcacagcg atcatctggc ttggaggcaa cgtcgaaacc ttctccactc ctgaccat 2880
caaaaagaag gggagacaca ttaaaaccag aatgctggtc gccacatcc cgtcaccat 2940
ccttggcaga gactatctcc aggcctggg cgttaaacct gtgctggcac aactgtctaa 3000
ggaaatcaag ttccgcaaga tcyagctgaa agagggcaca atgggtlcaa aaalcccaca 3060
gtggccctcg accaaagaga agcttgaggg cgttaaggaa atcgtgagc gccctgttcc 3120
tgaggccaa gattagcgag ccagcgacaa taaccttac aacagcccca tctttgtgat 3180
taagaaagag agcggaacat ggagactcct gcaggacctg agggaaacta acaagacctg 3240
ccaggtcgga actgagatct ctccgggaut cctcacccc gggcgcttga ttaaatgcaa 3300
gcacatgaca gtctctgaca ttggagacgc ttattttacc atccccctg atcctgaatt 3360
tcgcccctat actgctttta ccatccccag calcaatcac caggagcccg ataaacgela 3420
tgtgtggaag tgcctccccc agggatttgt gcttagcccc tacatttaac agaagacact 3480
tcaagagalc ctcaacctt tccgcgaag ataccacag gtccaactct accaatatat 3540
ggacgacctg tcatggggt ccaacgggtc taagaagcag cacaaggaa ccatcatcga 3600
actgagggca atcctctctg agaaaggctt cgagacaccc gacgacaagc tgcagaaggt 3660
tctccatat agctggctgt gttaccagct ttgcctgaa aactggaaag tccagaagat 3720
gcagttggat atggtcaaga acccaacact gaacgacglt cagaagctca tgggcaatat 3780
tacctggatg agctccgga tccctgggct taccgttaa gacattgcc caactacaaa 3840
aggatgcctg gattgaacc agaaggtcat ttggacagag gaagctcaga aggaactgga 3900
ggagaataat gaaaagatta agaattgctc agggctccaa tactacaate cgaagaagta 3960
aatgttgtgc gaggtcgaaa tcaactaagaa ctactagtc accatlgta tcaaacagtc 4020
ccaaggctc ttgtgggccc gaaagaaaat catgaaggcc aacaaaggct gggtccacgt 4080
laaaaaatct atgtcctgc tccagcacgt cyccaccgag tctatcccc cgttcggcaa 4140
gtgccccacc ttcaagttc ccttcaacta ggagcagggt atgtgggaga tgcaaaaagg 4200
ctggtactac tcttggcttc ccagatcgt ctacacccc caagtggtg acgacagctg 4260
gagaatgaag cttgtcgagg agcccactag cgggaattaca atctatccc acggcggaag 4320
gcacaaacgga gagggaaatc ctgcatacgt cacatctaac ggcgcacca agcaaaagag 4380
gctcgccct ctactacacc aggtgggtga gaggatggct atccagatgg ccttgaggga 4440
cactagagac aagcagtgca aatltgtgac tgcagctac tactgtgga aaaacatcac 4500
agaggccct ggcttgagg gaccccagtc tccctgggtg cctatcatcc agaatactcg 4560
cgaaaaggaa atgtctatc tgcgctgggt gcctggacac aaaggaattt acggcaacca 4620
actcgccgat gaagccgcca aaattaaaga ggaatcatg ctgtccatcc agggcaaca 4680
galtaagay aagagagac aggcagctgg ctttgacctg tgtgtccat acgacatcat 4740
gatcccggt aggcacacaa agatcattcc aaccgatgtc aagatccag tgcaccccaa 4800
ttcatttggg tgggtgaccy gaaagtcacg catggctaag cagggtcttc tgattaaagg 4860
gggaatcatt gatgaaggat acaccggcga aatccagggt atctgcacaa atatgggcaa 4920
aagcaatatt agctcatc agggcgagaa gtctcctcaa ctcatcatcc tccagacca 4980
cagcaattca agacaacctt gggcagaaaa caagattagc cagagaggtg acaagggtct 5040
cggcagncga ggtgtgtct cctccagaa catccaggaa agcagagaa 5100
ttggcacacc tccctcaaga ttttggcccg caattacaag atcccactga ctgtgctaa 5160
gcagatcaca caggaatgcc cccactgcac caaacaaggt tctggccccc cggctgcgt 5220

```

gatgagggtcc cccaatcact ggccaggcaga ttgcacccac ctccgacaca aaattatcct 5280
gacctctgtg gagagcaatt ccggtctacat ccaagcaaca ctccctccca aggaaatgca 5340
atttgtgccc tcctctcgaa ttcttgaatg ggccaggctg ttctctccaa aatccctgca 5400
cacccacac ccacccaact ttgtggctga acctgtgtgt aatclgtctga agttccctgaa 5460
aatcgccccc accactggca ttccctatca ccttgaaage cagggtcattg tccagagggcg 5520
caacagaaat ctgaaagaaa agatccaatc tcacagagac aatacacaga cattggaggcg 5580
cgcacttcag ctgcgcccta tcaactgcaa caaaggaaaga gaaagcatgg cggcccgagac 5640
cccctgggag gtcttcacaa ctanccaggc ccaggtcate catgaaaagc tgcctcttga 5700
gcaggccccg tccctcaaaa agttctgctt ttataagatc cccggtgagc acgactggaa 5760
aggctctaca agagttttgt ggaaaggaga cggcgcaagt gtggtgaaag atgaggggca 5820
ggggatcact gctgtgcccc tgacacgcac caagcttctc atcaagccaa actgaaacgg 5880
acgaatccca gggggaatct caacccttat tacccaacag tcagaaaaat ctaagtgtga 5940
ggagaacaca atgtttcaac ctatgttta taataatgac agtaagaaca gcatggcaga 6000
atcgaaggaa gcaagagacc aagaaatgaa cctgaagaa gaattctaag aagaaaaaag 6060
aagaatgac tggctgaaaa taggtatgtt tctgttatgc ttacccaggg cccctctgaa 6120
ggtgaccagt ggtgcaagggt cctccggcag tctgttacctg aagaaaaaat tccatcaca 6180
acatgcatcg cgaagaagac cctgggacca gcccacacac aacatacacc tagcaggcgt 6240
gaccggtgga tcaggggaca aatactacaa gcagaagtac tccaggaaag actggaatgg 6300
aagaatcagag gagtacaaca ggccggccaaa gagctgggtg aagtaaatcg aggcatttgg 6360
agagcctat atttccgaga agaccaaaag ggagatttct cagcctgggg cggctatcaa 6420
cgagcacaag aacggcctct gggggaaacaa tctccaccaa gggtccttag acctggagat 6480
tcgaagcga tttatgactg ttgcatataa gcccaagaa gaactctcgc 6540
tatcccttgc tftggatttc ccttatggct attttggggg tggggggggc cgcttccctt 6600
tagtgagggt taatgctctg agcagacatg ataaagataa ttgatgagtt tggacaaaco 6660
acaactagaa tgcagtgaaa aaaatgcttt atttgtgaaa ttgtgtatgc tattgtctta 6720
tttgtaacca ttataagctg caataaacaa gttaaacaca acaattgcat tcattttatg 6780
tttcaggttc agggggagat gtgggagggt ttttaaagca agtaaaacct ctacaaatgt 6840
ggtaaaatc gataagatc gatccgggtc ggctgaatag cgaagaggcc cgcaccgatc 6900
gcccctccca acagtgcgc agcctgaatg gcgaatggac gggccctgta cggcgccatt 6960
aagcggggg ggtgtgtgtg ttacggcag cgtgaccgct accttgcca gcgcccctagc 7020
gcccctctt tccgctctt tccctctctt tctgcacag tctgcgggt tcccccgtca 7080
agctctaaat cgggggctcc ctttaggggt ccgatttaga gctttaccgc acctcgaccg 7140
caaaaaact gatttgggtg atggttcaag tagtgggcca tgcgctgat agacggtttt 7200
tcgcccctt agtttgggtg cccaggttct taatagtgga ctcttcttcc aaactggaa 7260
aacactcaac cctatctcgg cctatctcgg tctattcttt tgattataaa gggattttgc cgaatttgcg 7320
ctattgggtt aaaaatgagc tgatttaaca aatatttaac gcgaatttta accaaaatatt 7380
aacgtttaca atttgcctg atcggtatt ttctccttac gcatctgtgc ggtatttca 7440
accgcatatc cggatctcgc cagcaccatg gctgaaata acctctgaaa gaggaaactg 7500
gttagtacc ttctgaggc gaaagaacca gctgtggaat gtgtgtcagt tagggtgtg 7560
aaagtcccca ggctccccag caggcagaag tatgcaaaag atgcatctca attagtccgc 7620
aacagggtg ggaagttccc caggctcccc agcaggcaga agtatgcaaa gcatgcatc 7680
caatagtag caacacatag tcccgcctcc aactccgccc atcccgcctc taactccgccc 7740
cagttccgccc cacttccgccc cccatggctg actaattttt tttatctatg cagaggccga 7800
ggccgctcgc gctctgagc tatccagaa gtatgagga ggcttttttg gaggcctagg 7860
cttttgcaaa aagcttgatt ctctgcacac aacagttctg aacttaaggc tagagccacc 7920
atgattgaac aagatgagc gacgcagggt tctccggcgc cttgggtgga gaggctattc 7980
ggctatgact gggcacaaca gacaaatggc tgcctctgat ccgcccgtgt cccgctgtca 8040
gcgcaggggc gcccgctctc ttttgtcaag accgacctgt ccggtgcctc gaatgaactg 8100
caggacggag cagccggctg atcgtggctg gccacgacgg cgttctcttg cgcagctgtg 8160
ctcgacgttg tcaactgaagc gggaacggac tggctgctat tgggcgaagt gccggggcag 8220
gatctcctgt catctcaact tgcctctgoc gagaagaat ccatcatggc tgatgcaatg 8280
cggcgccctg atacgtctga tccggctacc tgcctctgoc accaccaagc gaaacatcgc 8340
atcgagcgag cagctatcgc gatggaagcc ggtctgttgc atcaggatga tctggagca 8400
gagcatcagg ggctcgcgcc agccgaactg ttcgcccagg tcaaggccgc cagcccgac 8460
ggcgaggatc tctgtgtgac gctgtgttgc cgaatcaat ggtggaaat 8520
ggccgctttt ctggattcat cagctgtggc cggctgggtg tggcggaagg ctatcaggac 8580
atagcgttgg ctacccgtga tattgttgaa gagcttggcg gcgaatgggc tgacccgttc 8640
ctcgtgcttt acggatcgc cgtcccgat tccgagcga tccgcttcta tccgcttctt 8700
gacgagttct tctgagcggg actctgggtg tcgaatgac cgaaccaagc accgccaacc 8760
tgccatcagc atggccgcaa taaaatatct ttattttcat tacatctgtg tgttgggttt 8820
tttgtgaaat cgaatgcgat aaggatccgc gtatgggtga ctctcagtaa aatctgctct 8880
gatgcgcgat agttaagcca gccccgacac ccgccaacac ccgctgacgc gccctgacgg 8940
gcttctctgc tcccgccatc cgtttacaga caagctgtga ccgctccagg gagctgcatg 9000
tgtcagggt tttcacgctc atcacgaaa cgcgcgagac gaaagggcct cgtgatcgc 9060

ctatttttat aggttaatgt catgataata atggtttctt agacgtcagg tggcactttt 9120
cggggaaatg tgcgcgggaa ccttatttgt ttatttttct aaatacattc aaatatgtat 9180
ccgctcatga gacaaatacc ctgataaatg cttcaataat attgaaaaag gaagagtatg 9240
agtaattcaac atttccgtgt cgcccttatt cctttttttg cggcattttg ccttccgtgt 9300
tttgtccacc cagaacacgt ggtgaaagta aaagatgctg aagatcagtt ggggtcagca 9360
gtgggttaca tccaactgga tctcaacagc ggtaagatcc ttgagagttt tcccccagaa 9420
gaacgttttc caatgatgag cacttttaaa gttctgctat gtggcgcggt attatccctg 9480
attgacgccc ggcaagagca actcgttgc cgcatacact attctcagaa tgacttgggt 9540
gagtactcac cagtcacaga aaagcatctt accgatggca tgacagtaag agaattatgc 9600
agtgctgcca taacctagag tgataacact ggcggccaact tacttctgac aacgatcga 9660
ggaccgaagg agctaacccg ttttttgcac aacatggggg atcatgtaac tccgcttgat 9720
cgttggaagc cggagctgaa tgaagccata ccaaacgacg agcgtgacac cagcatgcct 9780
gtagcaatgg caacaacgtt gcgcaaaact ttaactggcg aactacttac tctagcttcc 9840
cggcaacaat taatagactg gatggaggcg gataaagttg caggacacact tctgcgctcg 9900
gcccttccgg ctggtggtt tattgtgat aaatctggag ccggtgagcg tgggtctcgc 9960
gggtatcatt cagcactggg gccagatggt aagccctccc gtatcgtagt tatctacagc 10020
accgggagtc agccaactat ggatgaacga aatagacaga tccgtgagat aggtgcctca 10080
ctgattaaag attggttaact gtcagaccaa gttactcat atatacttta gattgattta 10140
aaacttcaat ttttaattaa aaggatctag gtgaagatcc tttttgataa tctcatgacc 10200
aaaaatccct aacgtgatt ttogttccac tgagcgtcag accccgtaga aaagatcaaa 10260
ggatcttctt gagatccttt ttttctgcgc gtaactgctt gcttgcaaac aaaaaacca 10320
ccgctacacg ccgtggtttg tttgccggat caaagactac caactctttt tccgaaggtta 10380
actggttca gcagagcgca gatccaaaat actgtccttc tagtgtagcc gtagttaggc 10440
caccacttca acaactctgt agcaccgctt acatacctcg ctctgctaat cctgttacca 10500
gtggtgctg cagtgggga taagtctgt cttaaccgggt tggactcaag ccgatagtta 10560
ccggataaag cgcagcggtc gggctgaacg ggggggttct gcacacagcc cagcttgag 10620
cgaaacacct acaccgaact gagataccta cagcgtgagc tatgagaagc cgcacgctt 10680
cccgaaggga gaaaggcgga caggatcccg gtaagcggca gggcggaac eggagagcgc 10740
acgaggagc ttcaggggcg aaacgcctgg tatctttata gtcctgtcgg gtttcgccc 10800
ctctgacttg acgctcgatt tttgtgatgc tctcaggggg gggcgagcct atggaaaaac 10860
gccagcaacg cggccttttt accggttctg gccctttgct ggcccttttg tccataggct 10920
cgacagatct 10930

<211> 11131

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY4.0Z

<400> 9

```

ctaaattgta agcgttaata ttttggttaa attcgcgtta aatttttgtt aaatcagctc 60
attttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga 120
gatagggttg agtggtgttc cagtttggaa caagagtgca ctattaaaga acgtggactc 180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc 240
ctaatcaagt tttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag 300
ccccgattt agagcttgac ggggaaagcc aacctggctt atcgaaatta atacgaactc 360
ctatagggag accggcagat cttgaataat aaaatgtgtg tttgtccgaa atacgcgttt 420
tgagatttct gtcgcgactt aaattcatgt cgcgcgatag tgggtgttat cgcgcgatga 480
gatggcgata ttggaaaaat tgatatttga aaatatggca tattgaaat gtccgcgatg 540
tgagtttctg tgaactgat atcgccattt ttccaaaagt gatttttggg cataccgatg 600
atctggcgat agcgtttata tctgttacgg gggatggcga tagacgaact tgggtacttg 660
ggcgattctg tgtgtcgcaa atatcgcatg ttcgatatag gtgacagacg atatgagget 720
atatcgccga tagaggcgac atcaagctgg cacatggcca atgcatacg atctatacat 780
tgaatcaata ttggccaatta gccatattat tcatgtggtt tatagcataa atcaatattg 840
gctattggcc attgcatacg ttgtatccat atcgtaatat gtacatttat attggctcat 900
gtccacacatt accgcctagt tgacattgat tattgactag tatttaatag taatcaatta 960
cggggtcatt agttcatagc ccatatatgg agttcccggt tacataaact acggtaaatg 1020
gccgcctcgg ctgaccgcccc aacgaccccc gccattgac gtcaataatg acgtatgttc 1080
ccatagtaac gccaataggc actttccatt gacgtcaatg ggtggagtat ttacggtaaa 1140
ctgccacatt ggccagtacat caagtgtatc atatgccag tccgccccct attgacgtca 1200
atgacggtaa atggcccgcc tggcattatg ccagttacat gaccttacgg gactttctca 1260

```

```

cttggcagta catctacgta ttagtcatcg ctattaccat ggtgatgcgg ttttggcagt 1320
acacccatgg cgttggatag cggtttgact caccgggatt tccaagtctc caccctatg 1380
acgtcaatgg gagtttgttt tggcaccaaa atcaacggga ctttccaaaa tgtcgttaaca 1440
actgcgatcg ccgcgccgtg tgaogcaaat gggcggtagg cgtgtacggg gggaggtcta 1500
tataagcaga gctcgtttag tgaacccggc actcagatcc tgcggtctga gtcccttctc 1560
tgctgggctg aaaaaggcctt tgtaataaat ataattctct actcagtcct tgtctctagt 1620
ttgtctgttc gagatcccat agttggcgcc cgaacaggga cctgagaggg gcgcagaccc 1680
tacctgttga acctggctga tcttaggatc cccgggacag cagagggaga cttacagaag 1740
tcttctggag gtgttccctg ccagaacaca gggagacagg taagatggga gacctttga 1800
catggagcaa ggcgctcaag aagttagaga aggtgacggt acaagggtct cagaatata 1860
ctactggtaa ctglaattgg gcgtlaagtc tagtagactt atttcatgat accaacttly 1920
taaaagaaaa ggactgcgag ctgagggatg tcatccattt gctgggaagt gtaactcaga 1980
cycgtgcagg acagaaaaa gaggccttgg aaagaacatg gtgggcaatt tctgctgtaa 2040
agatgggctc ccagattaat aatgtagtag atggaaagcg atcattccag ctccaaagag 2100
cgaaatatga aagaagaact gctaataaaa agcagctcga gccctctgaa gaatatctct 2160
agaactagtg gatcccccgg gctgcaggag tggggaggca cgtatggcgc tttggtcgag 2220
cgggatccgg ccattagcca tattttcat tggttatata gcataaatca atattggcta 2280
ttggccatcg catacgttgt atccatatca taatatgtac atttatatg gtccatgtcc 2340
aacattacog caatgttgac atgtattatt gaactagttat taatagtaat caattacggg 2400
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttaocg taaatggccc 2460
gctggctgga ccgcctcaac acccccgcc attgacgtca ataargacgt atgttcccat 2520
agtaacgcca ataggagact tccattgacg tcaatggggt gagtatttac ggtaaactgc 2580
ccacttggca gtacatcaag tgtatcatat gccaaagtag cccctattg acgtcaatga 2640
cggtaaatgg ccgcctcggc attatgccca gtacatgacc ttatgggact ttccactattg 2700
gcagtacatc tacgtattag tcatcgctat taccatgggt atgcgggtttt ggcagtasat 2760
caatggcggt ggaatagcgt ttgactcagc gggatttcca agtctccacc ccattgacgt 2820
caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaacaaact 2880
cgccccattg accgcaatgg gcggtaggca tgtacggtgg gaggtctata taagcagagc 2940
tcgttttagt aaccgtcaga tgcgctggag acgcccacca cgctgttttg acctccatag 3000
aagacccgg gaccgatcca gccctccgcg ccccaagctt cagctgctcg aggatctcg 3060
gatccgggga atcccccatg ctccagatcc accatggggg atccccgtct tttcaaacgt 3120
cgtgactcgg aaaaacctgg cgttacccaa cttaactgcc ttgcagacaa tcccccttct 3180
gccagctggc gtaatagcca agagcccgcc accgatcgcc cttcccaaca gtgcgcagc 3240
ctgaatggcg aatggcgctt tgcctggttt ccggcaccag aagcgtgtgc ggaagctgg 3300
ctggagtgcg atcttctcga ggccgatact tgcgtcgtcc cctcaaatg gcagatgcac 3360
ggttacgatg gcgccatcta caaccaogta acctatccca ttacggtcaa tccgcggtt 3420
gttcccacgg agaatccgac gggttcttac tgcgtccatc ttaagtgtga tgaagctgg 3480
ctacagggaag gccagacgag aattattttt gatggcggtt actcggcggt tcatctgtgg 3540
tgcaacgggc gctgggtcgg ttacggccag gacagtcggt tgcgctctga atttgacctg 3600
agcgcatttt tacgcgcggg agaaaaaccg ctccggttga tgggtcgtcg tggagtgcac 3660
ggcagttatc tgaagatca ggaatgtgg cggatgagcg gcattttccg tgacgtctcg 3720
ttgctgcata aaccgactac acaaatcagc gatttccatg ttgccactcg ctttaatgat 3780
gatttcagcc gcgtgtactt atggcnagggt gttcagatgt cggcgaggtt cgtgactac 3840
ctacgggtaa cagtttcttt atggcnagggt gaaacgcagg tccgcagcgg atcgcgtcac actacgtctg 3900
ttcgcggtg aaattatcga tgagcctggt ggttatgcgg atcgcgtcac actacgtctg 4020
aacgtcgaaa acccgaaaact gtggagcgcc gaaatccaga atctctatcg tgggtggtt 4080
gaactgcaca ccgcgcagcg caogctgatt gaagcagaag cctgcgatgt cggtttccgc 4140
gaggtgcgga ttgaaaaatg tctgctgctg ctgaacggca agccgttgct gattogaggg 4200
gttaacgctc acgagcatca tctctgcat ggtcaggtca tggatgagca gacgatggtg 4260
caggatatcc tgcgtgatga gcaaaacacg tttaacggcg tgcgtgttcc gcatatccg 4320
aaccatccgc tgtggttacac gctgtcgac cgtacggccc tgtatgtggt ggtatgaacc 4380
aatattgaaa ccaacggcat ggtgccaatg aatcgtctga ccgatgatcc cgcctggcta 4440
ccggcgatga gcgaacggct aacgcgaatg gtgcagcgcg atcgtaatca cccgagtggt 4500
atcatctggt cgtcgggaaa tgaatcaggc caggcgcta atcacagcg cgtgtatcgc 4560
tggatcaaat ctgctgatcc ttcccgcgg gtgcagtatg aaggcggcgg agccgacacc 4620
acggccaccg atattatttg cccgatgtac gcgcgcgtgg atgaagacca gcccttccgc 4680
gctgtcgoga aatggtccat caaaaaatgg ctttgcgtac ctggagagac gcgcccgtg 4740
atccttttgc aatacgccca ccgatgggt aacagctctg cgggtttcgc taaatactgg 4800
caggcgttcc gtcagtatcc cegtttacag ggcggtctcg tctgggactg ggtggatcag 4860
tcgtgtatta aatatgatga aaacggcaac ccgtggtcgg cttacggcgg tgattttggc 4920
gatagccgga acgatgccca gttctgtatg aacggtctgg tctttcgoga ccgcagccg 4980
catccagcgc tgacggagcg aaaaaccagc cagcagtttt tccagttccg tttatccggg 5040
caaacctcgc aagtaccac cgaataacct ttccgtcata cgcataacga gctcctgcac 5100
tggatggtgg cgtcgtatgg taagccgctg gcaagcgggt aagtgcctct ggtgtcgtc

```

```

ccacaaggta aacagttgat tgaactgcct gaactacccg agccggagag cgcggggcaa 5160
ctctggctca cagtacgggt agtgcaaccg aacgcgaccc catggtcaga agccggggac 5220
atcagcgctt ggcagcagtg gcgtctggcg gaaaacctca gtgtgacgct ccccgccgcg 5280
tcccacgcga tcccgatctt gaccaccagc gaaatggatt ttgtcatcga gctgggtaat 5340
aagcgttggc aatttaaacg ccagtcaggc tttctttcac agatgtggat tggcgataaa 5400
aaacaactgc tgacgcggct cgcgatcag ttcacccgtg caccgctgga taacgacatt 5460
ggcgttaagt aagccgaccc cattgacctt aacgcctggg tgcagcgtg gaagcggcgg 5520
ggccattacc agcccgagac agcgttgttg cagtgcacgg cagatacaat tgcgtatgag 5580
gtgctgatta cgcacgctca cgcgtggcag catcagggga aaactttatt tatcagccgg 5640
aaaaactacc ggattgatgg tagtgttcaa atggcgatta ccgttgatgt tgaagtggcg 5700
agcgatacac gcgcatccggc gcggattggc ctgaactgcc agctggcgca ggtagcagag 5760
cgggttaaac ggctcggatt agggccgcga gaaaactatc ccgaccgctt tactgcgcgc 5820
tgttttgacc gctgggatct gccattgtca gacatgtata ccccgtaagt cttcccgagc 5880
gaaaaaggct tgcgtctggc gacgcgcgaa ttgaattatg gccacacca gtggcgcgcg 5940
gactttccagt tcaacatcag ccgctacagt caacagcaac tgatggaaac cagccalcgc 6000
catctgtctg acgcggaaga aggcacatgg ctgaataatg accggttcca tatggggatt 6060
ggtggcgagc actcctggag cccgtcagta tccgcggaat tccagctgag cgcgggtcgc 6120
taccattacc agtgtgtctg gtgtcaaaaa taataataac cgggcagggg ggatcccgag 6180
atccgctagt ggaatgtgtg tcagttaggg tgtggaagt cccagcctc cccagcagcg 6240
agaagtatgc aaagcatgcc tgcagggaat cgtatcaag cttatcgata ccgtccgact 6300
cgaaggcggg ccgcgtaccc agcttttgtt cccttagtg agggttaatt cgcggggaag 6360
tatttatcac taatcaagca caagtaatac atgagaaact ttactacag caagcacaat 6420
cctccaaaaa attttgtttt tacaaaaatc ctggtgaaca tgattggaag ggacctacta 6480
gggtgctgtg gaagggtgat ggtgcagtag tagttaatga tgaaggaaag ggaataattg 6540
ctgtaccatt aaccggactt aagttactaa taaaaccaa ttgagtattg ttgcaggag 6600
caagacccaa ctaccattgt cagctgtgtt tctgaggtc tctaggaatt gattacctcg 6660
atgtctcatc aaggaagag aataacaaa gactgaaggc aatccaacaa ggaagacac 6720
ctcaatattc gttataaggt ttgatatatg ggagtatttg gtaagggggt aacatggtca 6780
gcactgcact ctatggggga atcccagggg gaatctcaac cctattacc caacagtcag 6840
aaaaattcaa gttgtgagtg aacacaatgt ttcaacctta ttgttataat aatgacagta 6900
agaacggcat ggcagaatcg aagggaagcaa gagaccaaga aatgaacctg aaagaagaat 6960
ctaagaaga aaaaagaaga aatgactggg ggaataatag tatgtttctg ttatgcttag 7020
caggacatc tggaggaata ctttgggtgt atgaaggact cccacagcaa ctatatatag 7080
ggttggtgac gataggggga agattaaacg gatctggcca atcaaatgct atagaatgct 7140
ggggttcctt ccgggtgtgt agaccatttc aaaattactt cagttatgag accaatagaa 7200
gcattcatat ggataataat actgctacat tattagaagc tttaaccaat ataactgctc 7260
tat.aaataa aaaaacgaat tagaaccatg gaagttagta aagacttctg gcataactcc 7320
tttacctatt totttctgaag ctaacactgg actaattaga cataagagag attttgggat 7380
aagtgaata gtggcagcta ttgtagccgc tactgctatt gctgctagcg ctactatgct 7440
ttatgttctc ctaactgagg ttaacaaaat aatggaagta caaatcata cttttgaggt 7500
agaaaatagt actctaaatg gtatggattt aatagaacga caaataaaga tatttatatg 7560
fatgattctc caaacacatg cagatgttca actgttaaag gaaagacaac aggttagagga 7620
gacatttaat ttaattggat gtatagaag aacacatgta ttttgtcata ctggtcaccc 7680
ctggaatatg tcatggggac atttaaatga gtcaacacaa tgggatgact gggtaagcaa 7740
aatggaagat ttaaatcaag agatactaac tacacttnat ggagccagga acaatttggc 7800
acaatccatg ataacattca atacaccaga tagtatagct caatttggaa aagacctttg 7860
gagtcatatt ggaatattga ttctgtgatt gggagcttcc attataaat atatagtgat 7920
gtttttgtct atttatttgt tactaacctc ttgcctaaag atcctcaggg cctctggaa 7980
ggtagacagt ggtgcagggt cctccggcag tctgtacctg aagaaaaaat tccatcacia 8040
acatgcatg cgagaagaca cctgggacca ggcacaacac aacatacacc tagcaggcgt 8100
gaccggtgga tccaggggaca aatactacaa gcgaagtae tccaggaacg actggaatgg 8160
agaatcagag gagtacaaca ggcggccaaa gagctgggtg aagtcacac aggcatttgg 8220
agagagctat atttccgaga agaccaaaag ggagatttct cagcctgggg cggctatcaa 8280
cgagccacaag aacggcctctg gggggaacaa tctccaccaa gggctccttag acctggagat 8340
tcgaagcgaa ggaggaacaa tttatgactg ttgcattaaa gcccagaag gaaetctcgc 8400
tatcccttgc tatggatttc ccttatggct attttgggga ctagttaata tagtaggagc 8460
catagcaggc tatggattac gtggactcgc tgttataata aggatttgta ttagaggctt 8520
aaatttgata tttgaataaa tcagaaaaat gcttgattat atttgaagag ctttaaalcc 8580
tgccacatct catgtatcaa tgcctcagta tgtttagaaa aacaaggggg gaaetgtggg 8640
gtttttatga ggggttttat aaatgattat aagagtaaaa agaaagtgc tgatgctctc 8700
ataaccttgt ataacccaaa ggaactagctc atgttgcag gcaactaaac cgcataaacc 8760
gcatttttga cgcgagttcc ccattgttga cgcgttaact tctgtttttt acagtatata 8820
agtgcttcta tcttgacaat tgggcaactc gattctcgcg tctgagtooc tctctcgtg 8880
ggctgaaaag gcctttgtaa taatatataa tctctactca gtcctgtctc ctagtttgtc 8940

```

```

tgttcgagat cctacagagc tcatgccttg gcgtaatcat ggctcatagct gtttcctgtg 9000
tgaatttgtt atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa 9060
gccgggggtg cctaatgagt gagctaactc acattaattg cgttcgcctc actgcccgcct 9120
ttccagtcgg gaaacctgtc gtgcagctg cattaatgaa tcggccaacg ccgggggaga 9180
ggcgggtttg gtattggggg ctcttcgcgt tcctcgctca ctgactcgct gcgctcggtc 9240
gttcggctcg ggcgagcggt atcagctcac tcaaaggcgg taatacgggt atccacagaa 9300
tcaggggata acgcaggaaa gaacatgtga gcaaaaagcc agcaaaaagc caggaaacct 9360
aaaaaggcgg cgttgctggg gttttcccat agggctccgc cccctgacga gcatcccaaa 9420
aatcgacgct caagtacag gtggcgaaac ccgacaggac tataaagata ccaggcggtt 9480
ccccctggaa gctccctcgt gcgctctcct gttccgaccc tgccgcttac cggatacctg 9540
tcgcctcttc tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc 9600
agttcggtgt aggtcggtg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc 9660
gaccgctcgg ccttatccgg taactatcgt cttgagtgca acccggtaaq acacgactta 9720
tcgcccactg cagcagccac tggtaacagg attagcagag cgagggtatg aggcggtgct 9780
acagagttct tgaagtgggt gcctaactac ggctacacta gaaggacagt atttgggtac 9840
tgcgctctgc tgaagccagt taccttcgga aaaagagttg gttagctctt atccggcaaa 9900
caaaccaccg ctggtagcgg tgggtttttt gtttgcaagc agcagattac gcgcagaaaa 9960
aaaggtatcc aagaagatcc tttgatcttt tctacggggt ctgacgctca gtggaaagaa 10020
aacccacggt aagggtatct gggtcatgaga ttatcaaaaa ggatcttcac ctgatctctt 10080
ttaaatataa aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac 10140
agttaccaat gcttaatcag tgaggcaact atctcagcga tctgtctatt tcggtcatcc 10200
atagttgcct gactccccgt cgtgtagata actacgatac gggagggtctt accatctggc 10260
cccagtgctg caatgatacc gcgagaccca cgcctacccg ctccagattt atcagcaata 10320
aaccagccag ccggaagggc cgagcgccga agtggtcctg caactttatc cgcctccatc 10380
cagcttatta attgttgccg ggaagctaga gtaagtagtt cgccagttaa tagtttgccg 10440
aacgtttgtt ccattgctac aggcacgtg gtgtcacgct cgtcgtttgg tatggcttca 10500
ttcagctccg gttcccaacg atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa 10560
gcggttagct ccttcggctc tcgcatcgtt gtcagaagta agttggccgc agtgttatca 10620
ctcaggttla tcgagcactc gcataattct ctactgtca tgccatccgt aagatgcttt 10680
tctgtgactg gtgagtactc aaccaagtca tcttgagaat agtgtatgcy gcgaccgagt 10740
tgctcttgcc cggcgtaaat accgggataa accgcgccac atagcagaac tttaaaagt 10800
ctcaatattg gaaaacgttc ttccggggcg aacctctcaa ggatcttacc gctgttgaga 10860
tccagttcga tgaacccac tcgtgcaccc aactgatctt cagcatcttt tactttcacc 10920
agcgtttctg gctgagcaaa aacaggaagg caaaatgccg caaaaaaggg aataaggggc 10980
acacggaatg gttgaatact catactcttc ctttttcaat attattgaag catttatcag 11040
ggttattgtc tcatgagcgg atacatattt gaatgtattt agaaaaataa acaaataggg 11100
gttcgcgcga catttccccg aaaagtgcc c 11131

```

<210> 10

<211> 10998

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY8.0Z

<400> 10

```

agatcttgaa taataaaatg tgtgtttgtc cgaataacgc gttttgagat ttctgtcgcc 60
gactaaatcc atgtcgccgc atagtgggtg ttatcgccga tagagatggc gatattggaa 120
aaattgatat ttgaaatat ggcatattga aaatgtcgcc gatgtgagtt tctgtgtaac 180
tgatatccgc atttttccaa aagtgtatgt tgggcatacg cgtatctcgg cgtatagcgt 240
tatatcgttt acgggggagt gcgatagacg actttggtga cttgggcgat tctgtgtgtc 300
gcaaatatcg cagtttcgat ataggtgaca gacgatatga ggctatatcg ccgatagagg 360
cgacatcaag ctggcacatg gccaatgcac atcgatctat acattgaatc aatattggcc 420
attagccata ttattcattg gttatatagc ataaatcaat attggctatt ggccatttga 480
tacgttgtat ccatatcgta atatgtacat ttatatggc tcatgtccaa cattaccgcc 540
atgttgacat tgattattga ctagtattta atagttaatc attacggggt cattagtcca 600
tagcccatat atggagttcc gcgttacata acttacggta aatggccgcg ctggctgacc 660
gcccaacgac ccccgcccat tgacgtcaat aatgacgtat gttcccatag taacgccaat 720
agggactttc catgacgtc aatgggtgga gtatttacgg taaactgcc acttggcagt 780
acatcaagtg tatcatatgc caagtcggcc cctattgac gtcaatgacg gtaaatggcc 840
cgctggcgt tatgcccagt acatgacctt acgggacttt cctacttggc agtacatcta 900
cgtattagtc atcgctatta ccatggtgat gcggttttgg cagtacacca atggcgctgg 960

```

```

atagcgggttt gactcacggg gatttccaag tctccacccc attgaagtcg atggggagttt 1020
gtttttggcac caaaatcaac gggactttcc aaaatgtcgt aacaactcgg atcgcccgcc 1080
ccgttgacgc aatggggcgg taggcgtgta cgggtggagg tctatataag cagagctcgt 1140
ttagtgaacc gggcactcag attctgcggt ctgagtcctc tctctgctgg gctgaaaagg 1200
cctttgtaat aaatataaatt ctctactcag tccctgtctc tagtttgtct gttcgagatc 1260
ctacagttgg cgcgcgaaca gggacctgag aggggcgcag accctacctg ttgaacctgg 1320
ctgatcgtag gatccccggg acagcagagg aqaacttaca gaagtctctc ggaggtgttc 1380
ctggccagaa cacaggagga caggtaagat tgggagaccc tttgacattg gagcaaggcg 1440
ctcaagaagt tagagaaggt gacggtacaa gggctcaga aattaactac tggtaactgt 1500
aattggcgcg taagtctagt agacttattt catgatacca actttgtaaa agaaaaaggac 1560
tggcagctga gggatgtcat tccattgctg gaagatglaa ctcagaogct gtcaggacaa 1620
gaaagagagg cctttgaaag aacatggtgg gcaatttctg ctgtaaaagat gggcctccag 1680
attaataatg tagtagatgg aaaggcatca ttccagctcc taagagcgaa atatgaaaag 1740
aagactgcta ataaaaagca gtctgagccc tctgaagaat atctctagaa ctagtggatc 1800
ccccggctcg caggagtggg gaggcacgat ggccgctttg gtccaggcgg atccggccat 1860
tagccatatt attcatgtgt tatatagcat aaatcaatat tggctattgg ccatgtcata 1920
cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaac ttacggccat 1980
gttgacattg attattgact agttattaat agtaataat tacggggtca ttagttcata 2040
gcccatatat ggaattccgc gttacataac ttacggtaaa tggcccgctt ggtcgaacgc 2100
ccaacacccc cggcccatlg acgtcaataa tgacgtatgt tcccatagta acgccaatag 2160
ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgccac ttggcagta 2220
atcaagtgtc tcatatgcca agtacgcccc ctattgacgt caatgacggt aaatggcccg 2280
cctgggatta tgcacagcac atgaccttat gggactttcc tacttggcag tacatctacg 2340
tattatctat cgctattacc atgggtgatgc ggttttggca gtacatcaat gggcggtgat 2400
agcgggtttg ctacacggga ttccaagtc tccaccccat tgacgtcaat gggagtttgt 2460
tttggcacca aaatcaacgg gactttccaa aatgtcgtaa caactccgct ccatgacgc 2520
aaatggcggg taggcattga cgggtggagg tctatataag cagagctcgt ttagtgaacc 2580
gtcagatcgc ctggagacgc catcacagct gttttgacct ccatagaaga caccgggacc 2640
gatccagctc cggcgccccc aagcttcagc tgcctgagga tctcggaatc cggggaattc 2700
cccagctcca ggaaccacca tgggggatcc cgtcgtttta caacgtcgtg actggaaaaa 2760
ccctggcggt acccaactta atcgcttgc agcacatccc cctttcgcca gctggcgtaa 2820
tagcgaagag gcccgcacgg atcgcccttc ccaacagttg cgcagcctga atggcgaatg 2880
gcgctttccc tggtttccgg caccagaagc ggtgccgaa agctggctgg agtgcgactc 2940
tccctggggc gatactgtcg tcgtcccttc aaactggcag atgcacgggt acgatgcgct 3000
catctcaccc aacgtaacct atccattac ggtcaatccg cgttttgttc ccacggagaa 3060
tccgaogggg tgttactcgc tcacatttaa tggttgatga agctggctac aggaaggcca 3120
gacgcgaatt atttttgatg gcgttaactc gggtttcat ctgtggtgca acgggcgctg 3180
ggtcggttac ggcacggaca gtctgttgc gtctgaattt gacctgagcg catttttacg 3240
cgccggagaa aaccgcctcg cgggtgatgt gctcggttgg agtgacggca gttatctgga 3300
agatcaggat atgtggcgga tgagcggcat tttccgtgac gtctcgttgc tgcataaacc 3360
gactacacaa atnagcgatt tccatgttgc cactcgcttt aatgatgatt tcagccgcgc 3420
tgtactggag gctgaatttc agatgtcggc cgagttgcgt gactaactac gggtaacagt 3480
ttcttttatg cagggtgaaa ccgaggtcgc cagcgggacc ggccttttgg ggggtgaaat 3540
tatcgatgag cgtggtggtt atgcgatcg cgtcacacta cgtctgaacg tcgaaaaccc 3600
gaaactgtgt agrgccgaaa tccgaatct ctactgtcgg gtggttgaac tgcacacgc 3660
cgacggcacg ctgattgaag cagaagcctg cgatgtcggg ttccgcgagg tgcggattga 3720
aaatggtctg ctgctgctga acggcaagcc gttcgtgatt cgaggcggtt accgtcacga 3780
gcatactctc ctgcatgttc aggtcatgga tgagcagacg atggtgcagg atatcctgct 3840
gatgaagcag aacaaattta acgcctgctg ctgttcgcat tatccgaacc atccgtgtg 3900
gtacacgctg tgcgacgcct acggcctgta tgtggtggat gaagccaata ttgaaaccca 3960
cggcatggtg ccaatgaatc gtctgaccca tgatccgcgc tggctaccgg cgatgagcga 4020
acgcgtaacg cgaatggtgc agcgcgatcg taatcacccc agtgtgatca tctggtcgtc 4080
ggggaatgaa tcaggccacg gcgctaatac cgacgcgctg tatcgtgga tcaaatctgt 4140
cgatccttcc cgcctggctg agtatgaagg cggcggaagc gacacacagg ccaccgatat 4200
tatttgcccg atgtacgcgc gcgtggatga agaccagccc ttcccgctg tgcgaaalg 4260
gtccatcaaa aaatggcttt cgcctacctg agagaocgcg ccgctgatcc tttgogaata 4320
cgccacgcgc atgggttaaca gctttggcgg tttcgctaaa tactggcagg cgtttcgtca 4380
gtatccccgt ttacagggcg gcttcgtctg ggaactgggt gatcagtcgc tgattaaata 4440
tगतgaaaao ggcaacccgt ggtcggctta cggcggtgat tttggcgata cgcggaacga 4500
tcgcaggttc tgtatgaac gtctggtctt tgcgcacccc acgcccatac cagcgtgac 4560
ggaaagcaaaa caccagcagc agtttttcca gttccgttta tccgggcaaa ccatcgaaat 4620
gaccagcgaa tacctgttcc gtcatagcga taacgagctc ctgcaactgga tgggtggcgt 4680
ggatggtaa cgcctggcaa gcggtgaagt gcctctggat gtcgctccac aaggtaaaca 4740
gttgattgaa ctgcctgaac taccgcagcc ggagagcgcc gggcaactct ggcctacagt 4800

```

```

acgcgtagtg caaccgaacg cgaccgcatg gtcagaagcc gggcacatca gcgcctggca 4860
gcagtgggcg ctggcggaac accctcagtg gacgctcccc gccgcgtccc acgccatccc 4920
gcattctgacc accagcgaaa tggattttttg catcgagctg ggttaataagc gttggcaatt 4980
taaccggccag tcagggttttc ttccacagat gtggattggc gataaaaaac aactgctgac 5040
gccgctggcgc gacacagttca ccctgacacc gctggataac gacattggcg taagtgaagc 5100
gaccgcgatt gaccctaacg cctgggtcga acgctggaag gcggcgggccc attaccagc 5160
cgaagcagcg ttgttgtagt gcacggcaga tacacttgct gatcggtgac tgattacgac 5220
cgctcacgcg tggcagcatc aggggaaaaac cttattttatc agccggaaaa cctaccggat 5280
tgatggtagt ggtcaaatgg cgtattaccgt tgattgttaa gtggcgagcg atacaccgca 5340
tcggcgcgcg attggcctga actgcoagct ggcgacagta gcagagcggg taaactggct 5400
cggattaggg ccgcaagaaa actatcccga ccgccttact gccgcctggt ttgaccgctg 5460
ggatctgcca ttgtacagca tgtatacccc gtacgtcttc ccgagcgaaa acggtctcg 5520
ctgcgggacg cgcgaattga attatggccc acaccagtg gcggcgact ccagttcaa 5580
catcagccgc tacagtcaac agcaactgat ggaaccagc catcgccatc tgtgcacgc 5640
ggaagaagcg acatggctga atatcgacgg ttcccatatg gggattggtg cgcacgactc 5700
clggagcccg tcagiatcgg cgyaaltcca gctgagcgcc ggtcgctacc attaccggtt 5760
ggtctggtgt caaaaataat aataaccggg caggggggat ccgacagatc ggctgtgaa 5820
tgtgtgtcag tttaggtgtg gaaagtcccc aggtctccca gcaggcagaa gtatgcgaag 5880
catgctcgca ggaattcgat atcaagctta tcgataccgt cgaactcgag gggcgcccg 5940
gtaccagct ttgtttccct ttagttaggg ttaattgcgc gggaagtatt tatcactaat 6000
caagcacag taatacctga gaaactttta ctacagcaag cacaactctc caaaaaattt 6060
tgtttttaca aaatccctgg tgaacatgat tggaaaggac ctactagggt gctgtggaag 6120
ggttagggcg cagtagtagt taatgatgaa ggaagggaa taattgctgt accattaac 6180
aggataagtg tactaataaa accaaattga gtattgttgc aggaagcaag acccaactac 6240
cattgtcagc tgtgtttcct qacctcaata tttgttataa ggttgcatat gaatcccagg 6300
gggaattcca accctctata cccaacagtc agaaaaatct aagtgtagg agaacaat 6360
gtttcaacct tattgttata ataatgcagc taagaacgc atggcgagaat cgaaggagc 6420
aagagcccaa gaatgaacct ctaaaagaa aaaaaagaa aatgtagctg gaatgtagc 6480
tggaaaaatg gtatgtttct gttatgctta gcaggaaacta ctggaggaa atcttggtag 6540
tatgaaggac tcccaacagc acattatata cgtatggggg aagattaaac 6600
ggatctggcc aatcaaatgc tatagaatgc tggggttccct tcccggggtg tagaccatt 6660
caaaattact tcagttatga gaccaataga agcatgcata tggataataa tactgctaca 6720
ttattagaag ctttaacctta tataactgct ctataaataa caaaacagaa tagaaaact 6780
ggaagttagt aaagactctc ggcataaact ctttaacctat ttctctgaa gctaacactg 6840
gactaettag acataagaga gatttttgta taagtgcact agtgtagccg 6900
ctactgctat tgtgtctagc gctactatgt cttatgttgc tctaactgag gttacaaaa 6960
taatggaagt acaaaalcal acttttgagg tagaaaaatag tactctaaat ggtalggat 7020
taatagaacg acaataaag atattatatg ctatgattct tcaaacacat cgagatgttc 7080
aactgttaaa ggaagacaa caggtagagg agacatttaa tttaattgga tgtatagaaa 7140
gaacacatgt attttgtcat actggtcatc cctggaatat gtcatgggga catttaaatg 7200
agtcaacaca atgggatgac tgggtaagca aaatggaaga tttaaatcaa gagatactaa 7260
ctacacttca tggagccagg aacaatttgg cacaatccat gataacattc aataccacag 7320
atagtatagc tcaatttggg aaagaccttt ggagtcatat tggaaattgg attcctggat 7380
tgggagcttc cattataaaa tatatagtga tglttttgct tatttatttg ttaactaacct 7440
cttcgctcaa gatcctcagg gccctctgga aggtgaccag tggtgacagg tctctccgca 7500
gtcgttacct gaagaaaaaa ttccatcaca aacatgcac gcgagaagac acctgggacc 7560
aggcccaaca caacatacac ctgacggcg tgacgggtg atcaggggac aaatactaca 7620
agcagaagta ctccaggaac gactggaatg gagaatcaga ggagtacaac aggcggccaa 7680
agagctgggt gaagtcaatc gaggcatttg gagagagcta tatctccag aagaccaag 7740
gggagatttc tcagcctggg cggctatca acgagcacia gaaaggctct ggggggaa 7800
atcctacca aggtctctta gacctggaga ttcaagcga aggaggaac atttatgact 7860
gttgcattaa agcccagaa ggaactctcg ctatcccttg ctgtggattt ccttatggc 7920
tattttgggg actagtaatt atagtaggac gcatagcagg ctatggatta cgtggactcg 7980
ctgttataat aaggatttgt attagaggct taaatttgat atttgaaata atcagaaaa 8040
tgettatta tattggaaga gctttaaatc ctggcacatc teatgtate atgcctcagt 8100
atgtttagaa aaacaagggg ggaactgtgg ggtttttatg aggggtttta taaatgatta 8160
taagagtaaa aagaaagtgt ctgatgctct cataaccttg tataacccaa aggaactagc 8220
catgttgcta ggcacataaa ccgcaataac cgcatttttg atcgcagttc cccattggg 8280
acgcgttaac ttctgttttt tacagtatat aagtgcctgt attctgacaa ttgggcactc 8340
agattctgcg gtcgtagtc ctctctctgct gggctgaaaa ggctttgta ataaatata 8400
ttctctactc agtccctgtc tctagtttgt ctgttcgaga tctacagag ctaatgcctt 8460
ggcgtlaatc tggtcatagc tgtttcctgt gtgaaattgt tatccgltca caattccaca 8520
caacatacga gccggaagca taaagtgtaa agcctggggg gcctaagag tagetaaat 8580
cacattaatt gcgttgcgt cactgcccgc tttccagtc ggaacctgt cgtgccagct 8640

```

```

gcattaatga atcgcccaac gcgcggggag aggcgggttg cgtattgggc gctcttccgc 8700
ttcctcgctc actgactcgc tgcgctcggg cgttcggctg cggcgagcgg taccagctca 8760
ctcaaggggc gtaatacggg tatccacaga atcaggggat aacgcaggaa agaaccatgtg 8820
agcaaaaggc cagcaaaagg ccagggaaccg taaaaaggcc gcgttgctgg cgtttttcca 8880
taggctccgc cccctcgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa 8940
ccgacagga ctataaagat accaggcgtt tccctctgga agctccctcg tggcgtctcc 9000
tgttccgacc ctgcgctta ccggatacct gtcgccttt ctcccttcgg gaagcgtggc 9060
gctttctcat agctcacgct gtaggtatct cagttcgggt taggtcggtc gctccaagct 9120
gggctgtgtg caggaacccc ccgttcagcc cgaccgctgc gctttatccg gtaactatcg 9180
tcttgagtc aaccgggtta gacacgactt atcgccactg gcagcagcca ctggtaacag 9240
gattagcaga gcgaggtatg taggcgggtg tacagagttc ttgaagtggg ggctaaacta 9300
cggctacact agaaggacag tatttgggat ctgcgctctg ctgaaggcag ttaccttcgg 9360
aaaaagagtt ggttagctct gatccggcaa acaaacccac gctggtagcg gtggtttttt 9420
tgtttgcaag cagcagatta ccgcagaaaa aaaaggatct caagaagatc ctttgatctt 9480
ttctacgggg tctgacgctc agtggaaacga aaactcacgt taagggaattt tggatcatgag 9540
attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat 9600
ctaaaglatc tatgagtaaa ctctggtctga cagttaccaaa tgcttaatac gtgaggcacc 9660
tatctcagcg atctgtctat ttctgtctac catagttgcc tgactccccg tctgttagat 9720
aactacgata cgggggggct taccatctgg ccccgctgct gcaatgatac cgcgagacc 9780
acgctcacgc gctccagatt tatcagcaat aaaccagcca gccggaaggc ccgagcgcac 9840
aagtgtctct gcaactttat ccgcctccat ccagctctatt aattgttgcc ggggaagctag 9900
agtaagtagt tgcgcaagta atagtttgcg caacgtttgt gccattgcta caggcatcgt 9960
gggtgtaacg tctcgctttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg 10020
agttacatga tccccatgt tgtgcaaaaa agcgggttagc tccttcggtc ctccgatcgt 10080
tgtcagaagt aagtgtggcg cagtggtatc actcatggtt atggcagcac tgataattc 10140
tcttactgtc atgcatccg taagatgctt ttctgtgact ggtgagtagt caaccaagt 10200
attctgagaa tagtgtatgc ggcgacggag ttgctcttgc ccggcgtaaa tacgggataa 10260
tacgcgcaca catagcaaaa cttaaaaagt gctcctcatt ggaaaaacgt ctccggggcg 10320
aaaaactcca aggatcttac cgtctgttag atccagttcg atgtaacca ctctggcacc 10380
caactgatct tcagcatctt ttaactttac cagcgtttct ggttgaggaa aaacaggaa 10440
gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaaac tcatactctt 10500
cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt 10560
tgaattgatt tagaaaaata aacaaatagg ggttcgcgcg acatttcccc gaagaagtgc 10620
acctaaattg taagcgttaa tatttgtta aaattcgcgt taaatttttg ttaaatcagc 10680
tcatttttta accaataggc cgaaatcggc aaaatccctt ataaatcaaa agaatagacc 10740
gagatagggt tgaagtgtgt tccagtttgg aacaaggagc cactattaaa gaacgtggac 10800
tccacgctca aaggggcgaaa aaccgtctat caggggcgatg gcccactacg tgaacatca 10860
ccctaataca gttttttggg gtcgaggtgc cgttaaagca taaatcgga cctaaagg 10920
agcccccgat ttagagcttg acgggggaaag ccaaccgggc ttatcgaaat taatacgact 10980
cactataggg agaccggc 10998

```

<210> 11

<211> 8870

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY8.1Z

<400> 11

```

agatcttgaa taataaaatg tgtgtttgtc cgaataacgc gttttgagat ttctgtcgcc 60
gaclaaaalc atgltcgugcg atagtggtyl ttatcgccga tagagalggc gatattggaa 120
aaattgatat ttgaaaaatg ggcatattga aaatgtcgcc gatgtgagtt tctgtgtaac 180
tgatctcgcc atttttccaa aagtgtattt tgggcatacg cgatatctgg cgatagcgct 240
tatatcgttt acgggggatg gcgatagacg actttgggtg cttgggcgat tctgtgtgtc 300
gcaaatatcg cagtttcgat ataggtgaca gacgatatga ggtctatatc ccgatagagg 360
cgacatcaag ctggcacatg gccaatgcat atcgatctat acattgaatc aatattggcc 420
attagccata ttattcattg gttatatago ataaatcaat attggctatt ggccattgca 480
tacgttgtat ccattatgta atatgtacat ttatattggc tcatgtccaa cattaccgcc 540
atgttgacat tgattattga ctagtattta atagttaatc attacggggt cattagtcca 600
tagcccatat atggagttcc gcgttacata acttaaggta aatggcccgc ctggctgacc 660
gccccacgac ccccgcccat tgacgtcaat aatgacgtat gttcccatag taacgccaat 720
agggactttc cattgacgtc aatgggtgga gtatttacgg taaactgccc acttggcagt 780

```

```

acatcaagtg tacatatgc caagtccgcc cctatttgac gtcaatgacg gtaaatggcc 840
cgccctggcat tacgccagtg acatgacott acgggacttt cctacttgge agtacatcta 900
cgtatagtc atcgctatta ccatgggtgat ggggttttgg cagtacacca atgggcgtgg 960
atagcgggttt gactcacagg gatttccaag tctccacccc attgacgtca atgggagttt 1020
gtttttggcac caaatcaaac gggactttcc aaaatgtcgt aacaactgcg atcgcccgcc 1080
ccgttgacgc aaatgggggg tagcggtgta cgtgggagg tctatataag cagagctcgt 1140
ttagtgaacc gggcactcag attctgcggg ctgagtcctc tctctgctgg gctgaaaagg 1200
cctttgtaat aaataataat ctctactcag tccctgtctc tagtttgtct gtccgagatc 1260
ctacagttgg cgcgccgaac gggacotgag agggggcgag accctacctg ttgaacctgg 1320
ctgatcgtag ga:ccccggg acagnagagg agaactfana gaagtctct: ggaggtgttc 1380
ctggccagaa cacaggaggga caggtaagat tgggagaccc ttgacattg gagcaaggcg 1440
ctcaagaagt tagagaaggt gacggtacaa ggggtctcaga aattaactac tggtaactgt 1500
aattggggcg taagtctagt agacttattt catgatacca actttgtaaa agaaaaggac 1560
tggcagctga gggatgtcat tcatttgcgt gaagatgtaa ctgacagcgt gtcaggacaa 1620
gaaagagagg cccttgaaag aacatgggtg gcaattttctg ctgtaaaagt gggcctccag 1680
attaataatg tagtagatgg aaaggcatca ttccagctcc taagacgcaa atatgaaaag 1740
aagactgcta ataaaaagca gtctgagccc tctgaagaat atctctagaa ctagtggatc 1800
ccccgggctg caggagtggg gaggcacgat ggcgcgtttg gtcgaggcgg atccggccat 1860
tagccatagt atcatctggt tatatagcat aaatcaatat tggctattgg ccatgtcata 1920
cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaca ttaccgcat 1980
gttgacattg attattgact agttattaat agtaataat tacggggtca ttagttcata 2040
ggccatata ggagtctcgg gttacataac ttacggtaaa tggcccgctt ggtgacccg 2100
ccacagaccc ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag 2160
ggactttcca ttgagctcaa tgggtggagt atttacggta aactgcccac ttggcagtag 2220
atcaagtga tcatatgccca agtacgcccc ctattgacgt caatgacggt aaatggcccg 2280
cctggcatca tgcccagta acacatttat gggactttcc tacttggcag tacatctacg 2340
tatttagtcat cgttatatcc atggtgatgc ggttttggca gtacatcaat gggcgtggat 2400
agcggtttga ctcaacggga ttccaagtc tccaccccat tgacgtcaat gggagtttgt 2460
tttggcgcca aaatcaagg gactttccaa aatgtctaa caactcgccc ccattgacgc 2520
aaatggggcg taggcagtga cgtgggagg tctatataag cagagctcgt ttagtgaaac 2580
gtcagatcgc ctggagagcg catccacgct gttttgacct ccatagaaga caccgggacc 2640
gatccagctt ccgcccggcc aagcttcagc tgcctgagga totgaggatc cggggaaatc 2700
ccagctctca ggaatccaca tgggggatcc cgtcgtttta caacgtcgtg actgggaaaa 2760
ccctggcgtt acccaactta atcgcttgc agcacatccc ccttccgcca gctggcgtaa 2820
tagcaagag gcccgacccg atcgcccttc ccaacagtgg ccagcctga atggcgaatg 2880
gcgcttttgc tggtttccgg caccagaagc ggtgcgggaa agctggctgg agtgcgatct 2940
tccggaggcc gactactgct tcttcccttc aaactggcag atgcaagggt acgatgcgcc 3000
catctacacc aacgtaacct atccattac ggtcaatccg ccgtttgttc ccacggagaa 3060
tccgacgggt tgttaetgct tcacatttaa tgttgatgaa agctggctac aggaaggcca 3120
gacgcgaatt atttttgat cgtttaactc ggcgtttcat ctgtggtgca accggcgctg 3180
ggtcgtttac ggcacggaca gtctgttgc gtctgaattt gacctgagcg caattttacg 3240
cgcgggagaa aaccgctcgt cgtgatggt gctgcgttgg agtgacggca gttatctgga 3300
agatcgggat atgtggggga tgagrggat tttccgtgac gtcteghtgc tgnataaacc 3360
gactacacaa atcagcgalt tccatgttgc cactcgcttt aatgatgat tcagcccgcc 3420
tgtactggag gtfgaagtte agatgtgngg caggttgcgt gactacntac gggtaacagt 3480
ttctttatgg cagggtgaaa cgcaggctgc cagcggcacc gcgctttctg gcggtgaaat 3540
tatcgttag cgtggtggtt atgccgatcg cgtcacacta cgtctgaacg tccgaaaacc 3600
gaaactgtgg agcgcgcaaa tccgaatct ctatcgtcgg tgggtgaaac tgcacacccg 3660
cgaaggcagc ctgattgaag cagaagcctg cgtatcggt ttcccgagg tgggatttga 3720
aaatggctcg ctgctgctga acggcaagcc gttgctgatt cgaggcgltta accgtcacga 3780
gcacatcct ctgcatggtc aggtcatgga tgagcagac atggttcagg atatcctgct 3840
gatgaagcag aacaacttta acgcccgtcg ctgttcgcat tatccgaacc atccgctgtg 3900
gtacacgctg tgcgaccgct acggcctgta tglgtggat gaagccaata ttgaaacca 3960
cggcatgggt ccaatgaatc gtctgaccga tgcctcgccg tggctaccgg cgtgagcga 4020
acgcgtaacg cgaatggtgc agcgcgatcg taalcaaccy agltgalca tctggctcgt 4080
ggggaatgaa tcaggccacg gcgctaata cagacgcctg tatcgctgga tcaaatctgt 4140
cgatccttcc cgcctcggtc agtatgaagg cggcggaagc gacaccaagg ccacogatat 4200
tatttgcggc atgtacgcgc gcgtggatga agaccagccc ttcccgctg tggcgaatg 4260
gtccatcaaa aaatggcttt cgtacactgg agagacgcgc ccgctgatcc tttgcaata 4320
cgccacggc atgggtaaac gtcttggcgg tttcgtcaaa tactggcagg cgtttcgtca 4380
gtatccccgt ttacaggggc gttcgtctc ggaactgggt gatcagtcgc tgattaaata 4440
tgatgaaaac ggcaaccctg ggtcggctta cggcggtgat tttggcgata cgcggaacca 4500
tcgccagttc tgtatgaag gtctggtctt tgcgacccg acgcgctgac 4560
ggaagcaaaa caccagcagc agtttttcca gttcgttta tccgggcaaa ccatcgaagt 4620

```

```

gaccagcgaa  taccctgttcc  gtcacatagcga  taacgagctc  ctgcactgga  tgggtggcgct  4680
ggatggttaag  ccgctgtgcaa  gcggtggaagt  gcctctggat  gtgcctccac  aaggtaaaca  4740
gttgattgaa  ctgcctgaaac  taccgcagcgc  ggagagcgcc  ggccaactct  gggtcacagt  4800
acgctagtg  caaccgaaacg  cgaccgcagtg  gtcagaagcc  ggccacatca  gcgctggca  4860
gcagtggtgt  ctggcggaac  acctcagtggt  gacgcctccc  gcgcgctccc  acgccaatccc  4920
gcactctgacc  accagcgaaa  tggatttttg  catcgagctg  ggttaataagc  gttggcaatt  4980
taaccgccac  tcaggctttc  ttccacagat  gtggattggc  gataaaaaac  aactgctgac  5040
ccgcgtgcgc  gatcagttca  ccgctgcacc  gctggataac  gacattggcg  taagtgaagc  5100
gaoccccat  gaccctaaag  cctgggtcga  acgctggaag  gcgcgcgcc  attaccagcg  5160
cgaagcagc  ttgttgcaqt  gcacggcaga  tacacttgct  gatcggtgc  tgattacgac  5220
cgctcacgc  tggcagcatc  aggggaaaac  cttatttate  agccggaaaa  cctaccggat  5280
tgatggtagt  ggtcaaatgg  cgattaccgt  tgatgttgaa  gtggcgagcg  atacaccgca  5340
tcggcgccgc  attggcctga  actgccagct  ggcgagtgta  gcagagcgcg  taacctggct  5400
cggtatagg  ccgcaagaaa  actatcccga  ccgccttact  gcgcgctgtt  ttgaccgctg  5460
ggactctgca  ttgtcagaca  tgtatacccc  gtacgtcttc  ccgagcgaaa  acggtctgcg  5520
ctgcgggagc  cgcgaattga  attatggccc  acaccagtgg  ccgcgcgact  tccagttcaa  5580
catcagcgcc  tacagtcaac  agcaactgat  ggaaccagc  catcgccatc  tctgcaacgc  5640
ggaagaggc  acatggctga  atatcgacgg  ttcccatatg  gggattgggt  gcgacgactc  5700
ctgcagccgc  tcagtatcgg  cggaaattcca  gctgagcgcc  qctcgctacc  attaccagtt  5760
ggtctggtgt  caaaaaaat  aataaccggg  caggggggat  ccgcagatcc  ggctgtggaa  5820
tgtgtgtcag  ttaggtgtg  gaaagtcccc  aggtccccca  gcaggcagaa  gtatgcaaac  5880
catgctgca  ggaattcgat  atcaagctta  tggataccgt  cgaattggaa  gagctttaa  5940
tcctggcaca  tctcatgtat  caatgcctca  gtatgtttag  aaaaacaagg  ggggaactgt  6000
gggtgtttta  tgggggttt  tataaatgat  tataagagta  anaagaaagt  tggatgtct  6060
ctcataacct  tgtatacccc  aaaggactag  ctcatgttgc  taggcaacta  aaccgcaata  6120
acgcatttg  tgacgcgagt  tccccattgg  tgacgcgtta  acttccgtgt  ttacagtat  6180
ataagtgct  gtattctgac  aattggggcac  tcagattctg  cggtctgagt  ccttctctg  6240
ctgggctgaa  aaggcctttg  taataaatat  aattctctac  tcagtcctct  tctctagttt  6300
gtctgttca  gatcctacag  agctcatgcc  ttggcgtaat  catggtcata  gctgtttct  6360
gtgcaaatc  gttatccgct  cacaactcca  cacaacatac  gagccggaag  cataaagtgt  6420
aaagcctggg  gtgcctaatg  agtgagctaa  ctccatttaa  ttgcgttgcg  ctccactgcc  6480
gctttccagt  cgggaaacct  gtctgcccag  ctgcattaat  gaatcgccca  acgcgcgggg  6540
agaggcggt  tgcgtattgg  gcgtcttcc  gcttctcgc  tccactgact  gcgcgctcg  6600
gtcttccgc  tgccggcagc  ggtatcagct  cactcaaaag  cggtataacg  gttatccaca  6660
gaacagggg  ataacgcagg  aaagaacatg  tgagcaaaa  gccagcaaaa  ggccaggaac  6720
cgtaaaaagg  ccgcgttgct  ggctgttttc  cataggtccc  gcccccctga  cgagcatcac  6780
aaaaatcgac  gctcaagtca  gaggtggcga  aaccgcagac  gactataaag  ataccaggcg  6840
tttccccct  gaagctccct  cgtgcctct  cctgttccga  ccctgcgct  taccggatac  6900
ctgtccgct  ttctcccttc  gggaagcgtg  gcgcttttc  atagctcacg  ctgtaggtat  6960
ctcagttcgg  tttaggtcgt  tgcctccaag  ctgggctgtg  tgacgcaacc  ccccgttcag  7020
cccgaacgct  gcgccttacc  cgttaactat  cgtcttgagt  ccaaccgggt  aagacagac  7080
ttatccccc  tggcagcagc  cactggtaac  aggtattaga  gagcgaggta  tgtaggcggt  7140
gctacagagt  tcttgaagt  gtggctaac  tacggctaca  ctagaaggac  agtatctgt  7200
alctgcgct  tgcgtgaagc  agtlaacct  ggaaaaagag  ttggtagctc  ttgatccgc  7260
aaaaaacc  ccgctggtag  ccggtgttt  ttgttttga  agcagcagat  tacgcgcaga  7320
aaaaaaggat  ctcaaggaga  tcccttgatc  ttctctacgg  ggtctgacgc  tcagtggaa  7380
gaaactcac  gttagggaat  ttgtgtcatg  agattatcaa  aaaggatctt  cacttagatc  7440
cttttaaat  aaaaatgaag  ttttaaatca  atctaaagta  tatatagata  aacttggtct  7500
gacagttacc  aatgcttaat  cagtgaagca  cctatctcag  cgtatctgt  atttcttca  7560
tccatagttg  cctgactccc  cgtcgttag  ataactacga  tacgggaggg  cttaccatct  7620
ggccccagtg  ctgcaatgat  accgcagac  ccaagctcac  cggtccaga  tttatcagca  7680
ataaacagc  cagcggaaag  ggccgagcgc  agaagtggct  ctgcaacttt  atccgcctcc  7740
atccagctca  ttaattgttg  ccgggaagct  agagtgaagta  gttccagct  taatagttg  7800
cgcaacgttg  ttgccattgc  tacaggcatc  gtggtgtcac  gctcgtcgt  tgggtatggct  7860
tcattcagct  ccggttccca  acgatcaagg  cgagttaacat  gatccccat  gttgtgcaaa  7920
aaagcggtta  gctccttccg  tccctcagac  gttgtcagaa  gtaagtggc  cgcagtgtta  7980
tcactcatgg  ttatggcagc  actgcataat  tctcttactg  tcatgccatc  cgtaagatgc  8040
ttttctgtga  ctgggtgagta  ctcaaccaag  tcatcttgag  aatagtgtat  gcggcgaccg  8100
agttgctct  gcccgcgctc  aatacgggat  aataccgcgc  cacatagcag  aactttaaaa  8160
gtgtcatca  ttggaaaaacg  ttcttcgggg  cgaaaactct  caaggatctt  accgctgttg  8220
agatccagtt  cgatgtaacc  cactcgtgca  cccaactgat  cttcagcate  ttttaacttc  8280
accagcggtt  ctgggtgagc  aaaaacagga  aggcataaat  ccgcaaaaaa  gggataaagg  8340
cgacacggc  aatgtgaat  actcatactc  ttcttttttc  aatattatg  aagcatttat  8400
cagggttatt  gctcatgag  cggatacata  ttggaatgta  tttagaaaaa  taaacaaaa  8460

```

```

gggttccgc  gcacatttcc  ccgaaaagt  ccacctaagt  tgtaagcgtt  aatattttgt  8520
taaaattcgc  gttaaatttt  tgttaaatca  gctcattttt  taaccaatag  gccgaaatcg  8580
gcaaaatccc  ttataaatca  aaagaataga  ccgagatagg  gttgagtggt  gttccagttt  8640
ggaacaagag  tccactatta  aagaacgtgg  actccaactg  caaaggcgga  aaaaccgtct  8700
atcaggcgga  tggcccaacta  cgtgaacct  caccctaact  aagttttttg  gggctgaggt  8760
ccgctaagc  actaaatcgg  aaccctaag  ggagccccg  atttagagct  tgacgggaa  8820
agccaacctg  gcttatcgaa  attaaataga  ctactatag  ggagaccggc  8870

```

<210> 12

<211> 12481

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY3.1

<400> 12

```

agatcttcaa tattggccat tagccatatt attcattggg tatatagcat aaatcaatat 60
tggctattgg ccattgcata cgttgatctc atatcataat atgtacattt atattggctc 120
atgtccaata tgaccgccat gttggcattg attattgact agttattaat agtaaatcaat 180
tacgggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 240
tgccccgctc ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 300
tcccatagta accccaatag ggaactttcca ttgacgtcaa tgggtggagt atttacggta 360
aacgcgccac ttggcagtao atcaagtgta tcatatgcca agtcggcccc ctattgacgt 420
caatgacggt aaatggcccg cctggcatta tgcccagtao atgacctaac gggactttcc 480
tacttggcag tacatctacg tattagtrcat cgttattacc atggtagatgc ggttttggca 540
gtacaccaat gggcgtggat agcggtttga ctacgggga tttccaaagt tccaccccat 600
tgacgtcaat gggagtttgt tttggcacca aaatcaacgg gactttccaa aatgtcgtaa 660
caactgcgat cgcgcgcgcc gttgacgcaa atgggcggta ggcgtgtacg gtggggagtc 720
tatataagca gagctcgttt agtgaacogt cagatcacta gaagctttat tgcggtagtt 780
tatcacagtt aaattgctaa cgcagtcagt gcttctgaca caacagtcct gaacttaagc 840
tgcagtgact ctcttaaggt agccttgccg aagttggtcg tgaggcaact ggcaggtaag 900
tatcaagggt acaagacagg ttttaaggaga ccaatagaaa ctgggcttgt cgagacagag 960
aagactcttg cgtttctgat aggcacctat tggcttact gacatccact ttgcctttct 1020
ctcacaggt gtccactccc agttcaatta cagctcttaa ggctagagta cttaatagca 1080
ctcactatag gctagcctcg aggtcgcagc tatcgccga acagggacct gagaggggcy 1140
cagaccctac ctgttgaacc tggctgatcg taggatcccc gggacagcag aggaggaact 1200
acagaagctc tctggagggt ttcctggcca gaacacagga ggcaggtaa gatgggagac 1260
cctttgacat ggagcaaggc gctcaagaag tttagagaag tgacggtaca agggctctcg 1320
aaattaacta ctgctaactg taattgggcy ctaagtctag tagacttatt tcatgatacc 1380
aacctttgaa aagaaaagga ctggcagctg agggatgtca ttccattgct ggaagatgta 1440
actcagacgc tgtcaggaca agaaaagag gcttttgaag gaacatggtg ggcaattttc 1500
gctgtaaaag tgggctccca gattaataat gtagtagatg gaaaggcatc attccagctc 1560
ctaagagcga aatatgaaaa gaagactgct aataaaaagc agtctgagcc ctctgaagaa 1620
tatccaatca ttagatagtg ggcgggaaac agaaatttta gacctctaac acctagagga 1680
tatactaact gggtgaatac catcacagaca aatggtctat taaatgaagc tagtcaaaac 1740
ttatttggga tattatcagt agactgtact tctgaagaaa tgaatgcatt tttggatgtg 1800
gtacctggcc aggcaggaca aaagcagata ttacttgatg caattgataa gatagcagat 1860
gattggggata atagacatcc attaccgaat gctccactgg tggcacccac acaagggcct 1920
attcccatga cagcaaggtt tattagaggt ttaggagtag cttagagaaag acagatggag 1980
cctgcttttg atcagtttag gcagacatat agacaatgga taatagaagc catgtcagaa 2040
ggcatcaaa tgatgatttg aaaacctaaa gctcaaaata ttaggcaagg agctaaaggaa 2100
ccttacccag aatttgtaga cagactatta tcccaataa aaagtgaagg acatccacaa 2160
gagatttcaa aattcttgac tgatacactg actattcaga acgcaaatga ggaatgtaga 2220
aatgctatga gacatttaag accagaggat acattagaag agaaaatgta tgcctgcaga 2280
gacattggaa ctacaaaaa aaagatgatg ttattggcaa aagcacttca gactggtctt 2340
gcggggccat ttaaagggtg agccttgaaa ggagggcac taaaggcagc acaaacatgt 2400
tataactgtg ggaagccagg acatttatct agtcaatgta gagcacctaa agtctgtttt 2460
aaatgtaaac agcctggaca tttctcaag caatgcagaa gtgttccaaa aaacgggaag 2520
caaggggctc aaggggagcc ccagaaacaa actttccga tacaacagaa gactcagcac 2580
aacaaatctg ttgtacaaga gactcctcag actcaaaatc tgtaccocga tctgagcgaa 2640
ataaaaaagg aatacaatgt caaggagaag gatcaagtag aggatctcaa cctggacagt 2700
ttgtgggagt aacataatat ctagagaaaa ggcctactac aatagtatta attaatgata 2760

```

```

ctcccttaaa tgtactgtta gacacaggag cagatacttc agtgttgact actgcacatt 2820
ataatagggt aaaaatataga gggagaaaat atcaaggagc gggaataata ggagtgggag 2880
gaaatgtgga aacattttct acgctgtgga ctataaagaa aaagggtaga cacattaaga 2940
caagaatgct agtggcagat attccagtgga ctattttggg acgagatatt ctccaggact 3000
taggtgcaaa attggttttg gcacagctct ccaaggaaat aaaatttaga aaaaatagat 3060
taaaaggagg ccaaatgggg ccaaaaattc ctcaatggcc actcactaag gaaaaactag 3120
aaggggccaa agagatagtc caaagactat tgcacagagg aaaaatatca gaagctatgt 3180
acaataatcc ttataattca cccatatttg taataaaaaa gaggtctggc aaatggaggt 3240
tattacaaga tctgagagaa ttaaacaaaa cagtacaagt aggaacggaa atatccagag 3300
gattgctcca cccgggagga ttaattaaat gtaaacacat gactgtatta gatattggag 3360
atgcataatt cactatcccc tttagatccag agtttagacc atatacagct ttcactattc 3420
cctccattaa tcatcaagaa ccagataaaa gatattgttg gaaatgttta ccacaaggat 3480
tcgtgttgag cccatataata tatcagaaaa cattacagga aatttttcaa ccttttaggg 3540
aaagatatcc tgaagtcaaa ttgtatcaat atatggatga ttgtttcatg ggaagtaagt 3600
gttctaaaaa acaacacaaa gagttaatca tagaattaag gccgatctta ctggaaaagg 3660
gtttttgagc accagatgat aaattacaag aagtgccacc ttatagcttg ctaggtttatc 3720
aacctttgtc tgaaaattgg aaagtacaaa aaatgcaatt agacatggta aagaatccaa 3780
cccttaatga tggcaaaaa ttaatgggga atataacatg gatgagctca gggatcccaa 3840
ggttgacagt aaaaacacat gcagctacta ctaaggqatg tttagagttg aatcaaaaag 3900
taatttggac ggaagggcca caaaaaggat tagaagaaaa taatgagaag attaaaaatg 3960
ctcaaggggt acaatttat aatccagaag aagaatgtt atgtgaggtt gaaattacaa 4020
aaaaattatg accactttat gttataaaac aalcacaagg aatcctatgg gcaggtaaaa 4080
agattatgaa ggcataaag ggaatggcaa cagtaaaaaa tttaatgtta ttgttgcaac 4140
atgtggcaac agaaatgatt actagagtag gaaaatgtcc aacgtttaag gtaccattta 4200
ccaaagagca agtaattggg gaaatgcaaa aaggatggta ttattctctg ctcccagaaa 4260
tagtatatac acatcaagta gttcatgatg attggagaat gaaatttggt gaagaacctt 4320
catcaggaat aacaatatcc actgatgggg gaaaacaaaa tggagaagga atagcagctt 4380
atgtgaccag taatggggga actaaacaga aaaggttagg acctgtcact catcaagttg 4440
ctgaagaagt ggaatatcaa atggcattag aggataccag agataaaaaa gtaaatatag 4500
taactcatag ttattattgt tggaaaaata ttacagaagg attaggttta gaaggaccac 4560
aaagtctctg gtggcctata atacaaaaata tacagaaaaa agagatagtt tattttgctt 4620
gggtacctgg tccaaaaggg atatatggta atcaattggc agatgaagcc gcaaaaaata 4680
aagaacaaat catgctatga taccaaagga cacaatttaa agagaaaaga gatgaagatg 4740
cagggtttga cttatgtgtt ccttatgaca tcatgatacc tgbatctgac acaaaaatca 4800
taccacagga tgtaaaaatt caagttcttc ctaatagctt tggatgggtc actgggaaat 4860
catcaatggc aaacacaggg ttattaatata atggaggaaat aattgatgaa ggatatacac 4920
gagaaatata agtgatatgt actaatattg gaaaaagtaa tattaaatta atagagggac 4980
aaaaatttgc acaattaat atactacagc atcactcaaa ttccagacag ccttgggatg 5040
aaaaataaat atctcagaga ggggataaag gatlttgaag tacaggagta ttctgggtag 5100
aaaaatatta ggaagcacaa gatgaacatg agaattggca tacatcacca aagatattgg 5160
caagaaatata taagatacca ttgactgtag caaaacagat aactcaagaa tgtcctcatt 5220
gcactaagca aggatcagga cctgcaggtt gtgtcatgag atctcctaat cattggcagg 5280
cagattgcac acatttggac aataagataa tatgtacttt tgtagagtca aattcaggat 5340
acatacatgc tacattattg tcaaaagaaa atgcattatg tacttcatgt gctattttag 5400
aatgggcaag attgttttca ccaagtcctc tacacacaga taacggcact aattttgtgg 5460
cagaaccagt tgtaaatttg ttgaagttcc taaagatagc acataccaca ggaataccat 5520
atcatccaga aagtcagggg attgtagaaa gggcaaatag gaccttgaaa gagaagatc 5580
aaagtcatag agacaacact caaacactgg aggcagcttt acaacttctc ctacttactt 5640
gtaacaaagg gagggaaagt atgggaggac agacaccatg ggaagtattt atcaactaat 5700
aagcacaaag aatacatgag aaacttttac tacagcaagg acaatcctcc aaaaaatttt 5760
gtttttacaa aatccctggt gaacatgatt ggaagggacc tactagggtg ctgtggaagg 5820
gtgattgggc agtagtagtt aatgatgaag gaaagggaaat aattgctgta ccattaacca 5880
ggactaagct actaataaaa ccaaattgag tattgttgca ggaagcaaga cccaactacc 5940
attgtcagct gtgtttcctg aggtctctag gaattgatta cctcgatgct tcatlaagga 6000
agaagataaa acaagactg aaggcaatcc aacaaggaaag acaacctcaa tatttgtaal 6060
aaggtttgat atatgggagt atttggtaaa ggggtaacat ggtcagcatc gcatctctatg 6120
ggggaatccc agggggaatc tcaaccctta ttacccaaca gtcagaaaaa tctaagtgtg 6180
aggagaacac aatgtttcaa ccttatgttt ataataatga cagtaaaaac agcatggcag 6240
aatcgaaagg agcaagagac caagaaatga acctgaaaga agaactcaaa gaagaaaaaa 6300
gaagaaatga ctggtggaaa ataggtatgt ttctgttatg cttagcagga actactggag 6360
gaatactttg gtggtagtaa ggactccac agcaacatta tatagggttg gtggcagatg 6420
ggggaagatt aaacgatct gcccaatcaa atgctataga atgctggggt tcttccccc 6480
gggttagacc atttcaaaat tacttcagtt atgagaccaa tagaacatg catatggata 6540
ataactatgc taactatta gaagctttaa ccaatataac tgctctataa ataacaaaac 6600

```

```

agaattagaa acatggaagt tagtaaaagac ttctggcata actcctttac ctattttcttc 6660
tgaagctaac acgggactaa tttagacataa gagagatttt ggtataaagt caatagtggc 6720
agctatttga gccgctactg ctattgtctgc tagcgctact atgtcttatg ttgctctaac 6780
lgagggtlaac aaaataatgg aagtlacaaaa tcalactttt gaggtagaaa atagtactct 6840
aaatggtagt gatttaaatg aacgcacaaat aaagatatta tatgctatga ttcttcaaac 6900
acatgcagat gttcaactgt taaaggaaaag acaacaggta gaggagacat ttaatttaat 6960
tggatgtata gaaagaacac atgtattttg tcatctgggt catccttggg atatgtcatg 7020
gggacattta aatgagtcaa caaatggga tgactgggta agcaaaatgg aagatttaaa 7080
tcaagagata ctaactacac ttcatggagc caggaacaat ttggcacaal ccatgataac 7140
attcaatata ccagatagta tagctcaatt tggaaaagac ctttggagtc atattggaaa 7200
ttggattcct ggattgggag cttccattat aaaaatata tgatgtttt tgcttattta 7260
tttgttacta acctcttcgc ctaagatcct cagggccctc tgggaaggta ccagtgggtg 7320
agggtcctcc ggcatcggtt acctgaagaa aaaaattccat cacaacatg catcgcgaga 7380
agacacctgg gaccgggcc aacacaacat acacctagca ggctgacgg gtggatcagg 7440
ggacaataac tacaagcaga agtactccag gaacgactgg aatggagaat cagaggagta 7500
caacaggcgg ccaagagctg ggtgaagtc aatcgaggca tttggagaga gctatatctc 7560
cgagaagacc aaaggagaga ttctcagcc tggggcggct atcaacgagc acaagaacgg 7620
ctctgggggg aaacatctc accaagggtc cttagacctg gagattcgaa gcgaaggagg 7680
aaacatttat gactgttcca ttaagccca agaaggaaat ctgctatcc cttgctgtgg 7740
atttccctta tggctatttt ggggactagt aattatagta ggaagcatag caggctatgg 7800
attacgtgga ctgctgttta taataaggat ttgtattaga ggcttaaat tgatatttga 7860
aataatcaga aaaatgcttg attatattgg aagagcttta aatcctggca catctcatgt 7920
atcaatgctc cagtatgttt agaaaaaaa ggggggaact gtgggtttt tatgagggtt 7980
tttataatg attataagag taaaaagaaa gtgtctgatg ctctcataac cttgtataac 8040
ccaaaggact agtcatgttt gctaggcaac taaccgcgaa taaccgcat ttgtgacgga 8100
gttcccatc ttgtagcgtg ggtacctcta gactcgaccc ggccggccgc ttccttttag 8160
tgagggttaa tgttctgagc agacatgata agatataatg atgagtttgg acaaacacaa 8220
actgaatgc agtgaataaa atgctttatt tgtgaattt gtgatgctat tgcctttatt 8280
gtaacattta taagctgcaa taacaaght aacaacaaca attgcatcca ttttatgttt 8340
cagggttcagg gggagatgtg ggaggttttt taaagcaagl aaaaactcta caaatgttgt 8400
aaaatcagar aaggatcgat cggggtggc gtaatagcga agaggcccg accgatggc 8460
cttcccaaca gttgagcagg ctgaatggcg aatggagcgg cctgttagcg gcgcattaa 8520
cggcgggggg gtgtgtgtta cggcgagcgt gacgcgtaca cttgccagcg ccttagcgcc 8580
cgtctcttcc gctttcttcc ctctctttct cgccacgttc gccgggttcc ccgtcaacg 8640
tctaaatcgg gggctcctct taggggttcg atttagagct ttacggnaac tcgacgcgaa 8700
aaaacttgat ttgggttgat gttaacgtag tgggccaatg cctgtataga cggtttttgc 8760
ccctttgacg ttggagtcca cgttctttta tagtggactc ttgttccaaa ctggaacac 8820
actcaacctc atctcgctct attcttttga tttataagg attttgcga tttcggccta 8880
ttgttataaa aatgagctga tttacaaat atttaacgc aattttaaca aaatattaac 8940
gtttacaatt tgcctgatg cggatatttc tcttaacgca tctgtcggtt atttcacacc 9000
gcatacggcg atctgcgag caccatggc tgaaataacc tctgaaagag gaacttgggt 9060
aggtaacctc tgaggcgga agaacacgct gtggaatgtg tgtcagttag ggtgtggaaa 9120
gtccccagcg tcccccagc gcagaagtat gcaaaagcat catctcaat agtcagcaac 9180
cagggttgga aagttccag gctccccagc aggcagaagt atgcaaaaga tgcatctcaa 9240
ttagtacgca accatagtc ccgccctaac tccgccatc ccgccctaa ctccgccag 9300
ttccgccat tctccgcccc atggctgact aatttttttt atttatcgag agcccgaggc 9360
cgctcggcc tctgagctat tccagaagta gtgaggaagg ttttttggag gccataggct 9420
ttgcaaaaag ctgatatctc ctgacacaa agtctcgaac ttaaggctag agccaccatg 9480
attgaacagg atggattgca ccgaggttct ccgcccgtt gggtggagag gctattcggc 9540
tatgactggg cacaacagac aatcggtgc tctgatgcc ccgtgttccg gctgacagc 9600
caggggcgcc cggttctttt tgtcaagacc gaactgtccg gtgcctgaa tgaaactgac 9660
gacgagggca cggcgctatc gtggctggcc acgacgggct ttcttgcgc agctgtgctc 9720
gactgtgta cgtgaacggg aagggactgg ctgctattgg gcgaagtgc ggggcaggat 9780
ctctctgcat ctcaacttgc tctgcggag aagatctcca tcatggctga tgcaatggg 9840
cggctgcatc cgttgcaccc ggtacctgc ccatcgacc accaagcga acatcgcatc 9900
gagcgagcac gtaactggat ggaagccggg cttgtcgatc aggatgac cggcgaagag 9960
catcaggggc tgcgcagacc cgaactgttc gccaggctca aggcgcgcat gcccgacggc 10020
gaggatctcg tctgaccca ttggcgatgcc tgcctgcgca atatcatggt ggaaaaatgg 10080
cgctttctg gattcctga ctgaggcggt ctggtgttgg cggacogeta tcaggacata 10140
gcgttgctca ccgtgatgc tgcgaagag cttggcgggc aatgggctga ccgttctctc 10200
gtgctttag ctatcgacgc tcccgattcg ccgacatcg ccttctatcg ccttcttgac 10260
gagttcttct gagcgggact ctgggggttc aaatgacgca ccaagcgagc ccaaacctgc 10320
catacagat cgcgcaataa aatatcttta ttttcaatc atctgtgtgt tgggtttttt 10380
tgtgaatcga tagcataagg gatcccgcta ttgtgactc tcagtacaat ctgctctgat 10440

```

```

gccgcatagt taagccagcc ccgacaccg ccaacaccg ctgacgcgc ctgacgggct 10500
tgtctgtccc cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt 10560
cagaggtttt caccgtcaat accgaaacgc gcgagacgaa agggcctcgt gatcagccta 10620
tttttatagg ttaatgtcat gataataatg gtttctttaga cgtcagggtg caacttttcg 10680
ggaaatgtgc gccgaacccc tatttgttta tttttctaaa tacattcaaa tatgtatcgg 10740
ctvatgagac aataaccccg ataaatgctt caataatatt gaaaaaggaa gagtatgagt 10800
attcaacatt tccgtgtcgc ccttatctcc ttttttgcgg cattttgctc tctgtttttt 10860
gctcaccagg aaacgctggt gaaagtaaaa gatgetgaag atcagttggg tgcacgagt 10920
ggttacatcg aactggatct caacagcggg aagatccttg agagttttcg ccccgaaaga 10980
cgttttccaa tgatgagcac ttttaagatt ctgctatgtg gcgcgggtatt atcccgattt 11040
gacgcggggc aagagcaact cggctgcggc atacactatt ctacagaatg cttggttgag 11100
tactcaccag tcaagaaaaa gcatctttag gatggcatga cagttaagaga attatgcagt 11160
gctgccataa ccatgagtga taacacttgc gccaaacttac ttctgacaaac gatcggagga 11220
ccgaaggagc taaccgcttt ttgtcacaa atgggggata atgtaactcg ccttgatcgt 11280
tgggaacccg agctgaatga agccatacca aacgacgagc gtaacaccac gttgcctgta 11340
gcaatggcaa caacgttgcg caaactatta actggcgaa ctaactactc agctctcccg 11400
caacaattaa tagactggat ggaggcggat aaagtgtcag gaccacttct gcgctcggcc 11460
cttcggctcg gctggtttat tgctgataaa tctggagccg gtgagcgtgg gctctcggtt 11520
atcattgcag cactggggcc agatggttaag cctcccgcta tctgtattat ctacacgagc 11580
gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg 11640
attaagcatg gctaactgtc agaccaagtt tactcatata tacttttagt tgatttaaaa 11700
cttcattttt aatttaaaag gatctagggt aagatccttt ttgataatct catgacaaaa 11760
atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaggaa 11820
tctcttttag atcctgtttt tctgcgcgta atctgtgctg tgcaaaaaaa aaaaccacgg 11880
ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact 11940
ggcttcagca gagcgagat accaaaactc gtcctcttag ttagccgta gttaggccac 12000
cacttcaaga actctgtagc accgcctaca tactcgtctc tgctaactct gttaccagtg 12060
gctgctgcca gtgcgataa ctgctgtctt accgggttgg actcaagacg atagttaccg 12120
gataaaggcg agcggctcgg tgaacggggg ggttctgtga cacagcccg cttggagcga 12180
acgacctaca ccgaactgag atacctacag cgtgagctat gaaaaagcgc cagccttccc 12240
gaaggagaaa agggcgacag gtaactcggt agcggcaggg tgggaacagg agagcgacg 12300
agggagcttc cgggggaaa cgctcggtat ctttatagtc ctgtcgggtt tgcacacctc 12360
tgacttgagc gtcgattttg gtgatgtctg tcaggggggg ggagcctatg gaaaaacgcc 12420
agcaacggcg cctttttac ghtcctggcc ttttctggc cttttgtcca catggctga 12480
c 12481

```

<210> 13

<211> 6395

<212> DNA

<213> Artificial Sequence

<220>

<223> pCIneoERev

<400> 13

```

tgaataataa aatgtgtgtt tgtccgaat acgcgttttg agatttctgt cgcgcactaa 60
attcatgtcg cgcgatagtg gtgtttatcg ccgatagaga tggcgatatt ggaaaaattg 120
atatttgaaa atatggcata ttgaaaatgt cgcgcgatgt agtttctgtg taactgatat 180
cgccattttt ccaaaagtga tttttggcca tacgcgatat ctggcgatag cgcttatatc 240
gtttacgggg gatggcgata gacgactttg gtgacttggg cgattctgtg tgtcgcaaat 300
atcgcagttt ccatatagtg gacagacgat atcaggctat atcgcgcata gaggcgacat 360
caagctggca catggccaat gcatalcgat ctatacattg aatcaatatt ggccattagc 420
catattatct attggttata tagcataaat caatattggc tattggccat tgcatacgtt 480
gtaccatata cgtaatatgt acatttatat tggctcatgt ccaacatcac cgccatgttg 540
acattgatta ttgactagtt attaatagta atcaattacg ggggtcatag ttcatagccc 600
atattaggag ttccgcgtta cataacttac ggtaaatggc cgcctcggct gaccgcccaa 660
cgaccccccg ccattgagct caataatcac gtatgttccc atagtaacgc caatagggagc 720
tttccattga cgtcaatggg tggagtattt acggtaaaat gccactcgg ccgtacatca 780
agtgatcatc atgcacaagtc cgcctccatc tgaagtcagt gacggtaaat ggccgcctg 840
gcattatgcc cagtacatga ccttacggga ctttccctac tggcagtaga tctacgtatt 900
agtcatcgct attaccattg tgatcggttt ttggcagtag accaatgggc gtggatagcg 960
gtttgactca cggggatttc caagtctcca ccccatlgac gtcaatggga gtttgttttg 1020
gcacaaaat caacgggact ttccaaaatg tctgaacaac tgcgatcgcc cgcccgctg 1080

```

```

acgcaaatgg cgggtaggcg tgtacggtgg gaggtctata taagcagagc tcgttttagtg 1140
aacgcctaga tcaactagaag ctttattgag gtatgttata acagttaaat tgctaacgca 1200
gtcagtgctt ctgacacaaac agtctcgaaac ttaagctgca gtgactctct taaggtagcc 1260
ttgcagaagt tggctgtgag gcaactgggca ggtaagtata aaggttaaaa gacaggttta 1320
aggagaccaa tagaaactgg gcttgtcgag acagagaaga ctcttgcgtt tctgataggg 1380
acctattggt cttactgaca tccactttgc ctttctctcc acaggtgtcc actccagtt 1440
caattacagc tcttaaggct agagtactta atacgactca ctataggcta gtaacggccg 1500
ccagtgctgt ggaattcggc ttatggcaga atcgaaaggaa gcaagagacc aagaatgaa 1560
cctgaagaaa gaatctaaaag aagaaaaaag aagaaatgac tgggtgaaaa tagatcctca 1620
gggcccctcg gaaggtgacc agtgggtcgag ggtcctccgg cagtcgttac ctgaagaaaa 1680
aattccatga caaacatgca tgcgagaaag acacctggga ccaggcccaa caaacatcac 1740
acctagcagg cgtgacccgtt ggtacggggg acaataacta caagcagaag taactccagg 1800
acgaatggaa tggagaatca gaggagtaca acagcgggc aaagagctgg gtgaagtcaa 1860
tcgaggcatt tggagagagc tatatttccg agaagaccaa aggggagatt tctcagcctg 1920
gggcggttat caacgagcac aagaacggct ctggggggaa caatcctcac caagggtcct 1980
tagactgga gattcgaaag gaaggaggaa acatttatga agccgaalte tgcagatata 2040
catcacacgc ggggcgcgtt cccctttagt aggggttaat cttcgagcag acatgataag 2100
atacattgat ggttttgagc aaaccacaaac tagaatgcag tgaaaaaaat gctttatttg 2160
tgnaatttgt gatgtctatg ctttatttgt aaccattata agctgcaata aacagtttaa 2220
caacacaaat tgcattcaat ttatgtttca ggttcagggg gagatgtggg aggtttttta 2280
aagcaagtaa aacctctaca aatgtggtaa aatccgataa ggatcgatcc gggctggcgt 2340
aatagcaaa aggcgcgac cgtacgcctt tcccaacagt tgcgcagcct yaattgcyaa 2400
tggagcgcgc ctgtagcggc gcaatgaagc cggcggtgtg ggtgtttaa cgcgcagctga 2460
cgcctacatc tgcgcgcgc ctacgcgcgc tctcttctcc tcttctctcg 2520
ccacgttcgc cggcttctcc cgtcaagctc taaatcgggg gctcccttta ggggtccgat 2580
ttagaccttt acggcacctc gacgcgcaaa aacttgatll ygglyatggt tcaagtagtg 2640
ggccatcgcc ctgatgacg gtttttgcgc ctttgaggtt ggagtcacag tctcttaata 2700
gtggactctt gttccaaact ggaacaaacac tcaacccctat ccgggtctat tcttttgatt 2760
tataacggat ttgcccgtat tggcctattt gggttaaaaa tgagctgatt taacaaatat 2820
llaacggaa ltttaacaaa alattaacgt ttacaatttc gcctgatgag gtattttctc 2880
cttaccatc tgcgtgggat ttcacacgcg atacgcggat ctgcgcagca ccatggcctg 2940
aaataacccc tgaagaggga acttgggttag gtaccttctg agggcgaaag aaccagctgt 3000
ggaaatgtgt tcaagttagg tgtggaaggt cccagcgctc cccagcagc agaagtagc 3060
aaagcctgca tctcaattag tcagcaacca ggttgggaaa gtcccagcgc tcccagcag 3120
gcgaaggtat gcaaacgcat catctcaatt agtcagcaac catagctccg ccctaactc 3180
cgcccatccc gcccttaact cgcgccaggt cgcgccatcc tccgcccat ggtgactaa 3240
ttttttttt ttatgcagag gccgagggcg cctcggcctc tgagctattc cagaagttagt 3300
gaggaggtt ttttgaggc ctaggctttt gcaaaaagct tgattctctc gacacaacag 3360
tctgcactt aaggctagag ccaccatgat tgaacaagat ggattgcagc caggttctcc 3420
ggcgcgttg gggagaggc tattcggcta tgactgggca caacagacaa tccgctgctc 3480
tgatgcgcgc gtgttcggcg tgcagcgca gggcgcccg gtctcttttg tcaagacga 3540
cctgtccgtt gccctgaatg aactgcagga cgaaggcagc cggctatcgt ggctggccac 3600
gacggcggtt ccttgcgcag ctgtgctcga cgttgttact gaagcgggaa gggactggct 3660
gctattgggc gaagtgcgg ggcaggatct cctgtcatct cacttgcctc ctgcagaa 3720
agtatccatc atggtcgatg caatgcggcg gctgcacag ctgtgacgg ctaactgcc 3780
attgcaccac caagcgaaac atcgcatcga gcgagcact actcggatgg aagcgggtct 3840
tgtcgatcag gatgatctgg acgaagagca tcaggggctc gcgcagccg aactgttcgc 3900
caggctcaag ggcgcgatcg ccgacggcga ggaatcgtc gtgacccat gcgatgcctg 3960
cttgccgaat atcatgttg gaaaatggcg cttttctgga ttcacgaat gtggccggct 4020
gggtgtggcg gacgcctatc aggcacatagc gttggctacc cgtgatattg ctgaagagct 4080
tggcgcgnaa tgggtgaccc gcttctcgtt gctttacggt atcgcgcctc cegattcgca 4140
gcgcacgcgc tttctatcgcc tttctgacga gttcttctga cggggactct ggggttcgaa 4200
atgaccgacc aagcgacgcc caacctgcaa tcacgatggc cgcataaaaa tatctttatt 4260
ttcatcatc ctgtgtgttg gttttttgtg tgaatcgata gcgataagga tccgcgtatg 4320
gtgcactctc agtacaatct gctctgatgc cgtatagtta agccagccc gacacccgc 4380
aacacccgct gacgvgccct gacgggcttg tctgtctccc gcatccgtct acagacaagc 4440
tgtgacgcgc tccgggagct gcagtgtca gaggttttca cgtctatcac cgaacgcgc 4500
gagcgaag ggctcgtga taccgctatt tttataggtt aatgtcatga taataatggt 4560
ttcttgaag tcagggtgca cttttcgggg aaatgtgcgc ggaaccccta tttgtttatt 4620
tttcaaatc cattcaaat tgtatccgct catgagacaa taacctgat aaatgctca 4680
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgcgc tttttccct 4740
ttttcgcca ttttgcctc ctgtttttgc tcaacccgaa acgctcggtg aagtaaaaga 4800
tgtgaagat cagttgggtg cacgagtggt ttacatcgaa ctggatctca acagcggtaa 4860
gatcttgag agttttcgcc ccgaagaaag ttttccaatg atgagcactt ttaaagttct 4920

```

```

gctatgtggc gcggtattat cccgtattga cgcggggcaa gagcaactcg gtcgcccgat 4980
acaattttct cagaatgact tggltgagta ctccaccagtc acagaaaaagc atcttacgga 5040
tggcatgaca gtaagaqaal tatgcagtgc tgccataacc atgagtgata acactcgccg 5100
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacacat 5160
gggggatcat gtaactcgcc ttgatcggtg ggaaccggag ctgaatgaag ccataccaaa 5220
cgacgagcgt gacaccacga tgccctgtagc aatggcaaca acgttgccga aactattaac 5280
tggcgcaacta cttactctag cttcccgcca acaattaata gactggatgg aggcggataa 5340
agttgcagga ccacttctgc gctcggccct tccggctggc tggtttatgt ctgataaatc 5400
tggagccggt gagcgtgggt ctccgggtat cattgcagca ctggggccag atggtaaagc 5460
ctcccgatct gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag 5520
acagatcgct gagatagggt cctcactgat taagcattgg taactgtcag accaagttaa 5580
ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga tctaggtgaa 5640
gatccctttt gataactca tgacaaaaat cccttaacgt gagttttcgt tccactgagc 5700
gtcagacccc gtagaaaaa tcaaaaggatc ttcttgagat ccttttttct tgcgctgaat 5760
ctgctgcttg caaacaacaa aaccaaccgt accagcggtg gtttgtttgc cggatcaaga 5820
gctaccaact ctttttcgga aggttaactgg cttcagcaga gcgcagatac caaatactgt 5880
ccttctagtg tagccgtagt taggccacca cttcaagaac tctgtagcac cgcctacata 5940
cctcgctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt cgtgtcttac 6000
cggttgggac tcaagacgat agttaccgga taaggcgagc cggtcgggct gaacgggggg 6060
ttcgtgcaca cagcccgact tggagcgaa gacctacac gaactgagat acctacagcg 6120
tgagctatga gaaagcgcca cgttcccgga agggagaaag gcggaagagt atccggtaa 6180
cggcaggggt ggaacagagg agcgcagcag ggagcttcca gggggaaacg cctggtatct 6240
ttatagttct tcggggtttc gccacctctg acttgagcgt cgaattttct gatgtctgtc 6300
agggggggcg agcctatgga aaaaagccag caaccgggc tttttacggt tcttgccctt 6360
ttgctggcct tttgtcaca tggtctgaca gatctt 6395

```

<210> 14

<211> 5961

<212> DNA

<213> Artificial Sequence

<220>

<223> pESYNREV

<400> 14

```

tcaaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
aatatgaccg ccattgttggc attgattatt gactagttaa taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
gcctggctga ccgcccacag acccccgcct attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaaactgc 360
ccaattggca gtacatcaag tgtatcatat gccaaagtcc cccctaatg acgtcaatga 420
cggtaaatgg ccgcctctgc attatgocca gtacatgacc ttacgggact ttccactctg 480
gcagtaacac tacgtattag tcacgtcat taccatgggt atgcggttct ggcagtacac 540
caatgggcgt gtagtagcgt ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatgggagt ttgttttggc accaaaatca acgggaattt ccaaaatgct gtaacaactg 660
cgatcccccg ccccgctgac gcaaatgggc ggtaggcggt tacgggtggga ggtctabata 720
agcagagctc gtttagtgaa ccgtcagatc actagaagct ttattgctgt agtttatcac 780
agttaaattg ctaacgcagt cagtgccttc gacacaacag tctcgaaact aagctgcagt 840
gactctctta aggtagcctl gcagaayttg gtcgtgaggt actgggcagg taagtatcaa 900
ggttacaaaga caggttttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgttct tgataggcac ctattggctc tactgacatc caotttgcct ttctctccac 1020
aggtgtccac tcccagttca attacagctc ttaaggctag agtacttaat acgactcact 1080
ataggctagc ctgcagaatt cgcaccatg gctgagagca aggaggccag ggatcaagag 1140
atgaacctca aggaagagag caaagaggag aagcgccgca acgactgggt gaagatcgac 1200
ccacaaggcc cctggagggg ggaccagtg tgccgcgtgc tgagacagtc cctgcccgag 1260
gagaagattc ctagccagac ctgcactgcc agaagacacc tcggcccccg tcccaccag 1320
cadacacctc ccagaaggga taggtggatt aggggccaaga ttttgcaagc ctaggtctct 1380
caagaaaggc tggaaatggg aattaggggc gtgcaacaag ccgctaaga gctgggagag 1440
gtgaatcgcg gcactctggg ggagctctac ttccgcgagg accagagggg cgatttctcc 1500
gcattggggg gctaccagag ggcacaagaa aggctgtggg gcgagcagag cagccccgc 1560
gtcttgaggc ccgggagctc caaaagacgc cgcaaacacc tgtgaagtcg acccgggcgg 1620
ccgtctccct ttagtggagg ttaatgcttc gagcagacat gataagaac attgatgagt 1680

```

```

ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg 1740
ctatttgcctt atttgaacac attataagct gcaataaaca agttaacaac aacaattgca 1800
ttcatTTTTat gtttcaaggtt caggggggaga tgggggaggt tttttaaagc aagtaaaacc 1860
tctacaaaatg tggtaaaatc cgataaggat cgatccgggc tggcgtaata gcgaagaggc 1920
ccgcacccgat cgcctctccc aacagtttgcg cagcctgaat ggcgaaatgga cgcgcctcgt 1980
aggcgcgcat taagcgcggc ggggtgtgtg gttacgcgca cgtgagccgc tacacttgcc 2040
aggcgccctag cgcgcgcctc tttcgttttc ttcccttcc tctcgcgcac gttcgcgcgc 2100
tttcccgctc aagctctaaa tcgggggctc cctttagggt tccgatttag agctttacgg 2160
cacotcgacc gcaaaaact tgatttgggt gatggttccac gtatggggcc atcgccctga 2220
tagacggttt ttgcctctt ccacgttgag tccacgttct ttaatagtgg actcttctc 2280
caaacctgaa caacactcaa cctatctcgc gtctattctt ttgatttata agggattttg 2340
ccgatttctgg cctatttgggt aaaaaatgag ctgattttaa aaatatttaa cgcgaatttt 2400
aacaataat taacgtttac aatttcgcct gatcggtat tttctcctta cgcactctgt 2460
cggtatttca caccgcacac ggggatctgc gcagcaccat ggcttgaat aacctctgaa 2520
agagggaact ggttaggtac cttctgaggg ggaagaaacc agctgtggaa tgtgtgtcag 2580
ttagggtgtg gaaagtcccc aggtctccca gcaggcagaa gtatgcaag catgcatctc 2640
aattagtcat caaccaggtc tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa 2700
agcatgcac tcaattagtgc agcaaccata gtcccgcgcc taactccgcc catcccgccc 2760
ctaaactccgc ccaattccgc ccaattccgc ccccatggct gaactaatttt ttttatttat 2820
gcagagcgcg aggcgcgcct ggctctgag ctattccaga agtagtgagg aggtcttttt 2880
ggaggcgctg gcttttgcaa aaagcttgat tcttctgaca caacagctct cgaacttaagg 2940
ctagagccac catgtattgaa caagatggat tgcacgcagg ttctccggcc gcttgggtgg 3000
agaggctatc cggctatgac tgggcacaa acacaaatcg ctgctctgat gccgcgctt 3060
tcgggctgtc agcgagggg cgcccggttc ttttgtcaa gaccgacctt tccgggtccc 3120
tgaatcaact gcaggcagag gcagcgcggc tatcgtggct ggccacgacg ggcgttctct 3180
gcgcgctgt gtccagagtt gtactgaag cgggaaggga ctggctgcta ttgggcgaag 3240
tgccggggca ggtatctcgt tcatctcacc ttgctcctgc cgagaaagta tccatcatgg 3300
ctgatccaat cggcggtgtg catcagcttg atccggctac ctgccattc gaccaccaag 3360
cgaaacatcg catcgagcga gcacgtactc ggtggaaagc cgttctgtc gatcaggatg 3420
atctggacga agagcatcag gggctcgcgc cagccgaact gttcccgagg tccaaggcgc 3480
gcacgcccga cggcgaggat ctgctcgtga cccatggcga tgcctgcttg ccgaatctca 3540
tggcggaaaa tggcgcttt tctggattca tgcactgtgg ccggctgggt gtggcggacc 3600
gctatcagga catagcgttg gctaccgttg atattgtga agagcttggc ggcgaaatggg 3660
ctcgccgctt cctcgtgtgt tacggtatcg ccgctcccca ttcgcagcgc atcgccctct 3720
atcgctctct tgaacggttc ttctgagcgg gactctgggg ttgaaatga ccgaccaagc 3780
gacgcccac ctgccatcac gatggcgcca ataaaaatc tttattttca ttacatctgt 3840
gtgttgggtt tttgtgtgaa tggatagcga taaggatcgg cgtatgggtg actctcagta 3900
caatctgctc tgatgcgcga tagttaagcc agcccgcaca ccgcacca caacgctgacg 3960
cgccctgacg gcttgtctg ctcccgcat ccgcttacag acaagctgtg accgtctcgc 4020
ggagctgcat gtgtcagagg ttttcccggt catcccgaa acgcgcgaga cgaaaggccc 4080
tcgtgatcag cctattttta taggttaatg tcatgataat aatggtttct tagacgtcag 4140
gtggcacttt tcggggaaat gtgcgcggaa cccctatttg tttatttttc taaatacat 4200
caaatatgta tccgctcatg agacaataac cctgataaat gnttcaataa tattgaaaa 4260
ggaaaggtat gagtattcaa catttccgtg tgcctcttat tccctttttt gggcattttt 4320
gccttctgt ttttgcctac ccagaaacgc tgggtaaagt aaaaagatgct gaagatcagt 4380
tgggtgcacg agtgggttac atcgaaactg atctcaacag cggtaagatc cttgagagtt 4440
ttcgcccgca agaactgttt ccaatgatga gcacttttaa agttctgcta tgtggcgagg 4500
tattatccgc tattgacgcc gggcaagagc aactcggctc cgcatacac tattetcaga 4560
atgacttgggt tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa 4620
gagaattatg cagtgtctgc ataaccatga gtgataaac tgcggccaac ttacttctga 4680
caacgatcgg aggaccgaag gacttaacgc cttttttgca caacatgggg gatcatgtaa 4740
ctcgcttga tctgtgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca 4800
ccacgatgac ttagcaaatg gcaacaacgt tgcgcaact attaaactggc gaactactta 4860
ctctagcttc ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac 4920
ttctgcgtc gccctctccg gctggctgggt ttattgtgta taaatctgga gccggtgagc 4980
gtgggtctcg cgttatcatt gacgactgg ggcagatgg taagccctcc cgtatcgtag 5040
ttatctacac gacggggagt caggcaacta tggatgaacg aaatagacag atcgtgaga 5100
taggtcctc actgattaa agcattgtaac tgcagacca agtttactca tatatacttt 5160
agattgattt aaaaactcat ttttaattta aaaggatcta ggtgaagatc ctttttgata 5220
atctcatgac caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccocgtag 5280
aaaagatcaa aggatcttct tgagatcctt tttttctgcg cgtaatctgc tgettgcata 5340
caaaaaaac accgctacca gcggtgggtt gtttgccgga tcaagagata ccaactcttt 5400
ttcggaaagt aactgccttc agcagagcgc agataccaaa taactgtcct ctagtgtagc 5460
cgtagttagg caaccactc aagaactctg tagcacgcgc tacatactc gctctgctaa 5520

```

```

tctgttacc agtggctgct gccagtggcg ataagtcgtg tcttacgggg ttggactcaa 5580
gacgatagtt accggataag gcgcagcggc cgggctgaac ggggggttcg tgcacacagc 5640
ccagcttgga gcgaacgaco tacaccgaac tgagatacct acagcgtgag ctatgagaaa 5700
gcgcacgctc tccggaaggc agaaaggcgg acaggtatcc ggtaagcggc agggctggaa 5760
cagygagagc cacyagggag cttccagggg gaaacgcctg gtatctttal agtccctgctg 5820
ggtttcgcca cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc 5880
tatggaaaaa gcgcagcaac gcggcctttt tacggttccct ggcccttttcg tggccttttg 5940
ctcacatggc tcgacagatc t 5961

```

<210> 15

<211> 4307

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimised HIV gag-pol

<400> 15

```

atgggcgccc ggcgagcgt gctgtcgggc ggcgagctgg accgctggga gaagatccgc 60
ctggcccccg ggcgcaaaaa gaagtacaag ctgaagcaca tegtgtgggc cagccgcgaa 120
ctggagcgct tgcgctgaa ccccgggctc ctggagacca ggcgaggggtg ccccgagatc 180
ctcgcccaac tgcagccag cctgcaaac ggagcgaggg agctgcgcag cctgtacaac 240
accgtggcca cgtgtactg cgtccaccag cgcatacga tcaaggatag gaaagaggcc 300
ctggataaaa tgcgaagaga acagaataag agcaaaaaa aggcccaaca ggcgcgcgcg 360
gacaccggac acagcaacca ggtcagccag aactacccca tegtgcagaa catccagggg 420
cagatgggtg accaggccat ctccccccgc acgctgaacg cctgggtgaa ggtgggtgaa 480
gagaaggctt tttagcccgga ggtgatacc atgttctcag cctgtcaga gggagccacc 540
ccccagatc tgaacccat gctcaacaca gtggggggac accaggccgc catgcagatg 600
ctgaaggaga ccatcaatga ggaggtgccc gaattgggac gtgtgcaacc ggtgcacgca 660
gggcccctcg caccgggcca gatgcgtgag ccaaggggct cagacatcgc cggaaacgat 720
agtacccttc aggaacagat cggctggatg accaacaacc caccatccc ggtgggagaa 780
atctacaaac gctggatcat cctgggcctg aacaagatcg tgcgcatgta tagccctacc 840
agcactcctg acatccgcga aggcgcgaag gaaccttct cgcactacgt ggcgggttc 900
tacaacaacg tccgcgcgca gcaggctagc caggagggtg agaactggat gaccgaaacc 960
ctgctggctc agaaccgcga cccggactgc aagacgatcc tgaaggccct gggcccgagc 1020
gctaccctag aggaatgat gaccgcctgt caggggatgg gcggaccgcg ccacaaggca 1080
cgctcctcgg ctgaggccat gagccagggt accaacctcg ctaccatcat gatgcagcgc 1140
ggcaacttct ggaaccacag caagatcgct aagtgcctca actgtggcaa agaaggccac 1200
acagcccgca actgcagggc ccttagggaa aagggtctgt ggaactg:gg aaaggaaagg 1260
caccaaatga aagattgtac tgagagacag gctaattttt taggggaagt ctggccttcc 1320
cacaagggaa ggcaggggaa ttttcttcag agcagaccag agcccaacag cccaccagaa 1380
gagagcttca ggtttgggga agagacaaac actcctctc agaagcagga gccgatagac 1440
aaggaaactg atccttttag ttcctcaga tcaactcttg gcagcgacc ctgctacaa 1500
taaaatatgg gggcgagctc aaggaggtct tectggacac cggagcagac gacaccgtgc 1560
tggaggagat gtgcttgcca ggccgctgga agccgaagat gatcggggga atcggcggtt 1620
tcacaaaggt gcgcccagat gaccagatcc tcactgaat ctgcggccac aaggctatcg 1680
gtaccgtgct ggtgggcccc acaccgtca acatcatcgg acgcaacctg ttgacgcaga 1740
tcggttgca cgtgaacttc ccttaggac ctatcgagac ggtaccgtg aagctgaagc 1800
ccgggatgga cggccggaag gtccattgac agaggagaa agcaaggcac atcaaggcac 1860
tgggtggagt ttgcacagag atggaaaagg aagggaatat ctccaaagt ctcccgagaa 1920
accctgacaa caccgcgggt ttgcgaatca agaagaagg ctgcagaaa tggcgcaagc 1980
tgggtgactt tgggtgactt aacaagcgca cgaagaact ctgggaggt ctggcgggca 2040
tcccgcaccc cgcagggtg aagaagaaga aatcctgtac cgtactggat gtgggtgatg 2100
cctacttct cgttccccct gacgaagact tcaggaaagt cactgccttc acaatccct 2160
cgatcaacaa cgagacaacg gggattcgat atcagtacaa cgtgctgcgc cagggtctga 2220
aaggctcttc agcaatcttc cagagtagca tgacaaaaat cctggagcct tcccgcaaac 2280
agaacccgca catcgtcatc tatcagtaca tggatgactt gtactgggg ctctgatctag 2340
agatagggga gcaccgcacc aagatcgagg agctgcgcca gcacctgtg aggtggggac 2400
tgaccacacc cgacaagaag caccagaagg agcctccct cctctggatg ggttacgagc 2460
tgcaccctga caaatggacc gtgcagccta tegtgtctgc agagaaaagc agctggactg 2520
tcaacgacat acagaagctg gtggggaagt tgaactgggc cagtcagatt tacccaggga 2580
ttaaggtgag gcagctgtgc aaactcctcc gcggaaccaa ggcactcaca gaggtgatcc 2640
ccctaaccga ggaggccgag ctggaactgg cagaaaaccg agagatccta aaggagcccg 2700

```

```

tgacccggct gtactatgac ccctccaagg acctgatcgc cgagatccag aagcaggggc 2760
aaggccagtg gactatcag atttaccagg agcccttcaa gaacctgaag accggcaagt 2820
accgccggat gaggggtgcc cactactaac acgtcaagca gctgacccag gccgtgcaga 2880
agatcaccac cgaaagcatc gtgatctggg gaaagactcc taagttcaag ctgcccatcc 2940
agaaggaaac ctgggaaacc tggtaggacg agtatggca ggcaccc:gg attcctgagt 3000
gggagttctg caacacccct cccctgggtg agctgtggta ccagctggag aaggagccca 3060
tagtggggcg cgaaacottc tacgtggatg gggccgctaa caggagagct aagctgggca 3120
aagccggata cgtcactaac cggggcagac agaaggttgt caccctcact gacaccacca 3180
accagaagac tgagctgcag gccatttacc tgcctttgca ggaactcgggc ctggagggtg 3240
acatcgtgac agactctcag tatgcctcgg gcattcttca agcccagcca gaccagagt 3300
agtcggagct ggtcaatcag atcatcgagc agctgatcaa gaaggaaaag gttctatctg 3360
cctgggtacc cgcccacaaa ggcattggcg gcaatgagca ggtcgacaag ctggtctcgg 3420
ctggcatcag gaaggtgcta ttctctggat gcactgacaa ggcaccaggc gagcacgaga 3480
aataccacag caactggcgg gccatggeta gcgacttcaa cctgccccct gtggtggcca 3540
aagagatcgt ggcagctgt gacaagtgtc agctcaaggg cgaagccatg catggccagg 3600
tggactgtag ccccgcatc tggcaactcg attgcacca tctggagggc aaggttatcc 3660
tggtagcctg ccatgtggcc agtggctaca tggaggccga ggtcattecc gccgaaacag 3720
ggcaggagac agcctacttc ctctgaagc tggcaggccg gtggccagtg aagaccatcc 3780
atactgacaa tggcagcaat ttccactagt ctacggttaa ggcgcctgc tgggtggcgg 3840
gaatcaagca ggaattcggg atccccaca atccccagag tcaggggctg gtcgagtcta 3900
tgaataagga gttaaagaag attatcgccc aggtcagaga tcaggctgag catctcaaga 3960
ccgcggtcca aatggcggtc ttcatccaca atttcaagg gaaggggggg attggggggt 4020
acagtgcggg ggagcggtat gtggacatca tgcgaccca catccagact aaggagctgc 4080
aaaagcagat taccaagatt cagaatttcc ggtctacta cagggcacgc agaaaatccc 4140
tctggaaagg cccagcgaag ctctcttga aggttgagg ggcagtagt atccaggata 4200
atagcgacat caagtggtg cccagaagaa aggcgaagat cattagggat tatggcaaac 4260
agatggcggg tgaigtattg gtggcgagca gacaggatga ggattag 4307

```

<210> 16

<211> 4658

<212> DNA

<213> Artificial Sequence

<220>

<223> Codon optimised EIAV gag-pol

<400> 16

```

atgggcgcatc ccctcacctg gtccaaagcc ctgaagaaac tggaaaaagt caccgttcag 60
ggtagccaaa agcttaccac agccaattgc aactgggcat tgtccctggg ggatcttttc 120
cacgacacta atttcgttaa gqagaagatc tggcaactca gagacgtgat cccctctctg 180
gaggacgtga cccaacattt gtctgggcag gagcggaag ctctcgagcg caccgtgttg 240
gccatcagcg cagtcacaaat ggggtgcgaa atcaacaaac tgggtgacgg taaagctagc 300
tttcaactgc tcgcgcgtaa gtacgagaag aaaaccgcga acaagaaaca atccgaacct 360
agcgaggagt acccaattat gatcgacggc gccggcaata ggaacttcgg cccactgact 420
cccaggggct ataccacctg ggtcaacacc atccagacaa acggactttt gaacgaagcc 480
tcccagaacc tgttcggcat cctgtctgtg gactgcacct ccgaagaaat gaatgctttt 540
ctcgacgtgg tgcaggaca ggctggacag aaacagatcc tgcctgaagc cattgacaag 600
atcgccgacg actgggataa tgcacacccc ctgcccacgc cccctctggg ggctcccca 660
caggggccta tccctatgac cgctagggtc attaggggac tgggggtgcc ccgcgaagcc 720
cagatggagc cagcatttga ccaatttagg cagacctaca gacagtggat catcgaagcc 780
atgagcgagg ggattaaagt catgatcgga aagcccaagg cacagaacat caggcagggg 840
gccaaagaa ccatccctga glltgcgcag aggtctctgt cccagattaa atccgaagcc 900
caccctcagg agatctccaa gttcttgaca gacacactga ctatccaaaa tgcacaaatg 960
gagtcagaaa acgcatagag gcacctcaga cctgaagata cctcggagga gaaaatgtac 1020
gcatgtcgcg acattggcac taccagcaaa aagatgatgc tgcctgccc a ggctctgcaa 1080
accggcctcg ctggtccatt caaaggagga gcaactgaag gaggtccatt gaaagctgca 1140
caaatgttt ataattgtg gaagccagga catttatcta gtcaatgtag agcactaaa 1200
gtctgtttta aatgtaaaaa ccctggacat ttctcaaaag aatgcagaag tgttccaaaa 1260
aacgggaagc aaggggctca agggaggccc cagaaacaaa ctttcccgat acaacagaag 1320
agtcagcaca acaaatctgt tgtacaagag actcctcaga ctcaaaatct gtaccagat 1380
ctgagcgaaa taaaaaagga atacaatgtc aaggagaagg atcaagtaga ggatctcaac 1440
ctggacagtt tgtggagta acatacaatc tggagaagag gccactacc atcgctctga 1500
tcaatgacac ccctcttaat gtgctgctgg acaccggagc cgacaccagc gttctcacta 1560

```

```

ctgctcacta taacagactg aaatacacag gaaggaaata ccagggcaca ggcatcatcg 1620
cgcttggagg caacgtcgaa accttttcca ctctgtcac catcaaaaag aaggggagac 1680
acattaaaa cagaattgct gtccgcgaca tcccgctcac catcctggc catcctggc 1740
tccaggaact gggcgctaaa ctctgtctgg cacaactgtc taaggaaatc aagttccgca 1800
agatcgagct gaagaggggc acaatgggtc caaaaatccc ccagtggccc ctgaccaaaag 1860
agaagcttga gggcgctaaa gaaatcgctc agcgctctgt ttctgagggc aagattagcg 1920
aggccagcga caataaccct tacacacgac ccactcttgg gattaaagaa aggagcgga 1980
aatggagact cctgaggaac ctgagggaa tcacacagac cgtccaggtc ggaactgaga 2040
tctctcgagg actgctcac cccggcgccc tgattaaatg caagcacatg acagtctctg 2100
acattggaga cgcttatttt accatccccc tcatctctga atttgcctcc tatactgctt 2160
ttaccatccc cagcatcaat caccaggagc ccgataaacg ctatgtgtgg aagtgcctcc 2220
cccagggatt tgtgcttagc cctcacattt accagaagac acttcaagag atcctccaac 2280
ctttccgcta aagataccca gaggttcaac tctaccaata tatggacgac ctgttcatgg 2340
ggctccacgg gtctaagaag cagcaacaag aactcatcat cgaactgagg gcaatcctcc 2400
tggagaaaag cttcagagca cccgacgaca agctgcaaga agttcctcca tatagctggc 2460
tgggttacca gctttgcccc gaaaactgga aagtcacaga gatgcagttg gatatggta 2520
agaaacccac actgaacgac gtccagaagc tcatgggcaa tattacttgg atgagctccg 2580
gaatccctgg gcttaccgtt aagcacattg ccgcaactac aaaaggtatg ctggagttga 2640
accagaaggt catttggaca gaggaagctc agaaggaact ggagagaat aatgaaaaga 2700
tlaagaatgc tcaagggctc caatactaca atcccgaaga agaatgttg tycgaggtcg 2760
aaatcactaa gaactacgaa gccacctatg tcatcaaaac gtcccaggc atcttbtggg 2820
ccggaagaaa aatcatgaa gccaacaaag gctggtccac cgttaaaaat ctgatgctcc 2880
tgctccagca cgtcgccacc gagtctatca cccgcgtcgg caagtgcctc acctcaaaag 2940
ttcccttca taaggagcag gtgatgtggg agatgcaaaa aggtcgttac tactcttggc 3000
ttcccgagat cgtctacacc caccagtggt tgcacgaaga ctggagaatg aagcttctcg 3060
aggagcccac tagcggaatt acaatctata ccgacggcgg aaagcaaaac ggagagggaa 3120
tcgctgcata cgtcacatct aacggccgca ccaagcaaaa gaggtcgcgc cctgtcactc 3180
accaggtggc tgagaggtg gctatccaga tggcccttga ggacactaga gacaagcagg 3240
tgaacattgt gactgacgc tactactgct ggaaaaacat cacagagggc cttggcctgg 3300
agggacccca gttccctgg tggcctatca tccagaatat ccgcgaaaag gaaattgtct 3360
atttcgctg ggtgccttga cacaaggaa tttacggcaa ccaactcgcc gatgaaagcc 3420
ccaaaattaa agaggaaatc atgcttgcct accagggcac acagattaa gagaagagag 3480
acgaggaagc tggctttgac ctgtgtgtgc catacgacat catgattccc gttagcgaca 3540
caaaagatcat tccaaccgat gtcaagatcc aggtgcaccc caattcaatt ggttgggtga 3600
ccggaaggtc cagcatggct aagcagggtc ttctgattaa cgggggaatc attgatgaag 3660
gatacacccg cgaatccag gtgatctgca caaatatcgg caaaagcaat attaaagctt 3720
tcgaagggca gaagtgcgt caactcatca tctctcagca ccacagcaat tcaagacaac 3780
cttgggacga aacaagatt agccagagag gtgacaaggg cttcggcagc acaggtgtgt 3840
tctgggtgga gaacatccag gaagcacagg acgagcaaga gaattggcac acctccctca 3900
agattttggc cagcaattac aagatcccac tgactgtggc taagcagatc acacaggaat 3960
gccccactg caccaaaaca ggttctggcc ccgcggctg cgtgatgagg tccccaatc 4020
actggcaggc agattgcacc cactcgaca acaaaattat cctgaccttc gtggagagca 4080
attccggcta catccacgca acactcctct ccaaggaaaa tgcattgtgc acctccctcg 4140
caattcttga atgggcccag ctgttctctc caaaatccct gcacaccgac aacggcaca 4200
actttgtggc tgaacctgtg gtgaatctgc tgaagttcct gaaaaatcgcc cacaccactg 4260
gcattcccta tcacctgtaa agccagggca ttgtcgagag ggccaaacaga actctgaaag 4320
aaaagatcca atctccaga gacaatacac agacattgga ggccgcaactt cagctcgccc 4380
ttatccctg caacaaaaga agagaaagca tggcgcgcca gacccccgg gaggtcttca 4440
tactaacca ggcccaggtc atccatgaaa agctgctctt gcagcaggcc cagtcctcca 4500
aaaagtctg cttttataag atccccgggt agcacgactg gaaaggtcct acaagagttt 4560
tgtggaaaag agacggcgca gttgtgttga acgatgagg caaggggac atcgctgtgc 4620
ccctgacag caccaagctt ctcatcaagc caaactga 4658

```

<210> 17

<211> 10392

<212> DNA

<213> Artificial Sequence

<220>

<223> pIRESlhylgESYNGP

<400> 17

aattcgccac catgggcgat cccctcacct ggtccaaagc cctgaagaaa ctggaaaaag 60

```

tcaccgttca gggtagcdaa aagottacca caggcaattg caactgggca ttgtccctgg 120
tggatctttt ccacgacact aatttcgtta agggagaaaga ttggcaactc agagacgtga 180
lccccctctt ggaggacgtg acccaaacat tctctgggca ggagcgcgaa gctttcgagc 240
gcacctgggtg ggccatcagc gcagtcaaaa tggggctgca aatcaacaac gtggttgagc 300
gtaaagctag ctttcaactg ctccgcgcta agtacgagaa gaaaaccgac aacaagaaac 360
aatccgaac tggcgaggag tcccaattta tgatcgacgg cgcgggcaat aggaacttcc 420
gcccaactgac tcccgaggggc tataccacct gggtoaacac catccagaca aacggacttt 480
tgaacgaagc ctcccaaac ctgtlccgca tccgtctgtt ggaatgcacc lccgaagaaa 540
tgaatgcttt tctcgacgtg gtgcccaggac aggcctggaca gaaacagatc ctgctcgatg 600
ccattgacaa gatcgccgac gactgggata atcgccaccc cctgccaaac gccctctcgg 660
tggctcccc acagggggct atccctatga cgcctagggt cattagggga ctgggggtgc 720
cccgcgaaag ccagatggag ccagcatttg accaatttag gcagacctac agacagtggg 780
tcatcgaaag catcgagcag gggattaaag tcatgattcg aaagcccaag gcacagaaac 840
tcaggcaggg ggccaaggaa ccataccctg agtttctgca caggctctct toccagatta 900
aatccgaag ccacctcag gagatctcca agttctctgc agacacactg actatccaaa 960
atgcaaatga agagtgcaga aacggccatg ggcacctcag acctgaagat acctggagg 1020
agaaaatgtg cgcattgcgc gacattggca ctaccaagca aaagatgatg ctgctcgcca 1080
aggtctctga aaccggcctg gctggtccat tcaaggagg agcactgaag gggggtccat 1140
tgaagctgc acaaacatgt tataattgtg ggaagccagg acatttatct agtcaatgta 1200
gagcacctaa agtctgtttt aaatgtaaac agcctggaca tttctcaag caatgcagaa 1260
gtgttccaaa aaacgggagc caaggggctc aaggggggcc ccagaaacaa actttccgga 1320
tacaacgagc cagtcagcac aacaaatctg ttgtacaaga gactcctcag actcaaaatc 1380
tgtaccaga tctgagcgaa ataaaaagg aatacaatgt caaggagag gatcaagtag 1440
aggatctcaa cctggacagt ttgtgggagt acatcacaa ctcgagaaga ggccactac 1500
catcgtccct atcaatgaca cccctcttaa tgtgctgctg gacaccgag ccagacaccg 1560
cgttctcact acctctcact ataacagact gaaatacaga ggaaggaaat accagggcac 1620
aggtcctcct gggttggag gcaacgtcga aaccttttcc actcctgta ccatacaaaa 1680
gaaggcgaga cagattaaaa ccagaatgct ggtgcgcgac atccccgtca ccctccttgg 1740
cagagacat ctccaggacc tgggcgctaa actcgtgctg gcaaaactgt ctaaggaaat 1800
caagtctcgc caagatcgagc tgaacaggg cacaatgggt ccaaaaatcc ccagtgggc 1860
cctgacccaa gagaagcttg agggcgctaa ggaatctg cagcgctg cttctgaggg 1920
caagatttag gggcccgagc acaataaccc ttacaacagc cccatctttg tgattaaaga 1980
aaggcgggc aaatggagac tctgcaagg cctgagggaa cctcaacaga cgttccaggt 2040
cggaactgag atctctcgcg gactgcctca ccccgggc ctgattaaat gcaagacat 2100
gacagtctct cacttggag acgtttatct taccatcccc ctcgatcctg aatttgcgcc 2160
ctatactgct tttaacctcc ccagcatcaa tccaccggag ccgataaac gctatgtgtg 2220
gaagtgcctc ccccgaggat ttgtgcttag cccctacatt taaccagaag cacttcaaga 2280
gatcctccaa cctttcccg aaagataccc agaggttcaa ctctaccaat atatggagca 2340
cctgttccat ggttccaacg ggttcaagaa gcagcacaag gaactcatca tcgaactgag 2400
ggcaatccct ctggagaaag gcttcagac acccgagcac aagctgcaag aagtctctcc 2460
atatagctc ctgggctacc agctttgccc tgaaaactgg aaagtccaga agatgcagtt 2520
ggatatggtc aagaacccaa cactgaacga cgtccagaag ctcatgggca atattacctg 2580
gatgagctcc ggaatccctg ggttaccctg taagcacatt gccgcaacta caaaaggatg 2640
cctgggcttg aaccagaagg tcatctggac agagggaagc cagaaggaa cggaggagaa 2700
faatgaaaag attaaqaatg ctcaagggtt ccaatactac aatcccgagg aagaatgtt 2760
gtgcgaggtc gaaatcacta agaacthaga agccactat gtcatcaaac agtcccaagg 2820
catctgttgg gccggaagaa aaatcatgaa ggccaacaaa ggtggtcca cgttaaaaa 2880
tctgatgctc ctgctccagc acgtcgccac cgagtctatc acccggtcgc gcaagtggcc 2940
cactctcaaa gttcccttca ctgaaggaca ggtgatgtgg gagatgcaaa aaggtctgta 3000
ctactcttgg cttcccgaga tctctacac ccaccaagtg gtgcaagacg actggagaat 3060
gaagcttgtc gaggagccca ctagcggaat taacaactat ancgagcgcg gaaagcaaaa 3120
cggagcgga atcgtcgcat acgtcacatc taacggccgc accaagcaaa agaggctcgg 3180
ccctgtcact caaccaggtg ctgagaggat ggctatccag atggcccttg agggacctag 3240
agacacgag gtgaacattg tgactgacag ctactacgct tggaaaaaca tcacagaggg 3300
ccttggctcg gagggaaccc agtctccctg gtggcttate atccagaata tcccgaaaa 3360
ggaaattgtc tatttgcctt ggtgctctgg acacaaagga atttarggca accaactcgc 3420
cgatgaagcc gccaaaatta aagaggaaat catgcttgcc taccagggca cacagattaa 3480
ggagagagag gacgaggagc ctggctttga cctgtgtgtg ccatacgaca tcatgattcc 3540
cgttagcgac acaagatca ttccaaacga tgtcaagatc caggtgccac coaattoatt 3600
tggttgggtg accggaagt ccagcatggc taagcaaggt ctctgatta acqgggaat 3660
cattgatgaa ggataacacg gcgaatatca ggtgatctgn acaaatatcg gcaaaagcaa 3720
tattaaagct atcgaagggc agaagltcgc tcaactcatc atcctccagc accacagcaa 3780
ttcaagacaa cctlgggagc aaaaacagat tagccagaga ggtgcaaggg gcttcggcag 3840
cacaggtgtg ttctgggtgg agaactcca ggaagcacag gacgagcacg agaattggca 3900

```

```

caccctccct aagattttgg ccgcgaatta caagatccca ctgactgtgg ctaagcagat 3960
cacacagga aagattttgg ccgcgaatta caagatccca ctgactgtgg ctaagcagat 4020
gtcccccaat cactggcagg cagatttgcac ccacctcgac aacaaaatta tccctgacct 4080
cgtggagagc aatccgggtc acatccagcg aacactcctc tocaaggaaa atgcatgttg 4140
caccctccct gcaatttctg aatgggcccag cctgttctct ccaaaaacc tgacacccga 4200
caacggcacc aacttttggg ctgaacctgt ggtgaatctg ctgaagctc tgaaaatcgc 4260
ccaccccaat ggcatttccct atccacctga aagccagggc atgtgcgaga gggcccaacg 4320
aacctcga aaaaagatcc aatctcacag agacaataca cagacattgg agggccgact 4380
tcagctcgcc ctatccacct gcaacaaagg aagagaaagg atggggggcc agaccccttg 4440
ggagggtctc atcactaacc agggcccagg catccatgaa aagctgctct tgacgagcgc 4500
ccagctcctc aaaaagttct gcttttataa gatcccggt gacgacgact ggaagggtcc 4560
tacaagagtt ttgtggaaag gaggcggcgc agttgtggtg aacgatgagg gcaaggggat 4620
catcgctgtg cccctgacac gacccaagct tctcatcaag ccaaaclgaa cccggggcgc 4680
ccgcactaga ggaatttgcg cctctcctc ccccccctc aacgttactg gccgaagcgc 4740
cttggataaa ggcgggtgtg tgtttgtcta tatgtgattt tccaccatct tgcgctcttt 4800
tggcaatgtg agggcccagg aacctggccc tgtcttcttg acgagcatte ctagggggtc 4860
ttcccctct gccaaggaa tgcaaggctc gttgaatgtc gtgaaggaa agttctctct 4920
ggaagctctc tgaagacaaa caagctctgt agccacctt tgacggcagg ggaaccccc 4980
acctggcgac aggtgacctc gcggccaaa gccactgta taagataacc ctgcaaggc 5040
ggcacaaacc cagtgccacg ttgtgagttg gatagttgtg gaaagagtca aatggctctc 5100
ctcaagctga gtcaacaagg ggcctgaagg tgcccagaa gtaccccat gtatgggaat 5160
cctatctggc cactcggttc acatgcttta catgtgttta gtcagggtta aaaaagctc 5220
agggccccc aacccagggg acgtgggttt cctttgaaa acacgatgat aagcttgcca 5280
ccaccccggt ccaaaagatg atagatccgg aaagcctgaa ctaccgcga cgtctgtcga 5340
gaagtttctg atcgaaaagt tcgacagcgt ctccgacctg atgcagctct cggagggcga 5400
agaatctctg gctttcagct tcgatgtagg agggcgtgga tatgtcctgc gggtaaatag 5460
ctgcgcgcat ggtttctaca aagatcgtta tgtttatcgg cactttgcat cggccgcgct 5520
cccgatctcg ccaggtcctg acattgggga attcagcgag agcctgacct attgcatctc 5580
ccgcgctgca cagggtgtca cgttgcaaga cctgctgaa accgaactgc ccgctgttct 5640
gcagccggca cgggagccca tggatgcgat cgttcggcc gatcttagcc agacgagcgc 5700
gttcggccca ttgcgacgc aaggaaatcg tcaatacact acatggcgtg atttcatatg 5760
ccgagattgt gatcccatg tgtaoactg gcaaaactgt atggacgaca ccgtcagctc 5820
gtcgttcgag caggtctcg atgagctgat gtttgggccc gaggactgcc ccgaagtcgc 5880
gcacctcgtg cagcgagatt tcggctccaa caatgtcctg accgacaatg gccgcataac 5940
agcgtctatt gactggagcg aggcgatgtt cggggattcc caatacagag tgcacaacat 6000
cttctctctg aggcgctggt tggcttgtat ggagcagcag accgctactc tcgagcggag 6060
gcacccggag ctgcaggat ttgcacggca ttcgatgal gcagcttggg ccgagggctg 6120
ccaaactctat cagagcttgg atcgtccgat ccggagccgg gactgtcggg cgtacacaaa tcccccgcag 6180
atgcgcgcga atcgtccgat atggctgtgt agaagtaact gccgatagt gaaaccgacg 6240
aagcggcgcc gtctggacgc ccaaggata gactagatgc cgaccgaaca agagctgatt 6300
cccccaact cgtccgaggg aactcgcgc agcctagcaa ggcaaatgag agagaacggc 6360
tcgagacgc ctcagccagc tctcgtccac agltcgtcaa gctcgtcgg ctgggttcgc 6420
cttaccgttg gtggcagagt tcaggccctt ctggattgtg ttggtccca gggcacgatt 6480
ggagggccgg tcgcaatgat tgatctgact gatcccgag attggagatc gcccccgtg 6540
gtcatccca cgcactcggg cttagagctc ctgactgtgc ccttggaagg tgcccaatcc 6600
cctgccgat ttgtttgccc ctcccctgt gcatcgcat ttctgagtag gtylcaltct 6660
cagccatctt cctaataaaa tgaggaaalt gcatcgcat atttggaaga caatagcagg 6720
actgtccttt cctaataaaa tgaggaaalt gcatcgcat atttggaaga caatagcagg 6780
attctggggg gtgggggtgg gcaggacagc aagggggagg attggaaga ctggggctcg 6840
catgtctggg atgcggtggg ctctatgctc tctgaggcgg aaagaaccag ctggggctcg 6900
agtcatctct agttgtggtt tctccaaact catcaatgta tcttatcatg tctgtatacc 6960
gtcgaactct agctagagct tggcgtaate ctgttctctg ctgtttctct tgtgaaattg 7020
ttatccgctc acaattccac acaacatag agccggaagc ataaagtga aagcctgggg 7080
tgccataatg gtgagctaac tcacattaat tgcgttgcgc tccactgcgc ctttccagtc 7140
gggaaaacct tegtgcagc tgcattaatg aatccgcca ccccgcgagg agggcggttt 7200
gcgtattggg cgtctctccg cttectcgtc cactgactcg ctgcgctcgg tegtccgct 7260
gcggcgagcg gtatcagctc actcaaaagg ggtaatccgg ttatccacag aatcagggga 7320
taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggacc gtaaaaaggc 7380
cgcgttgcg gcgtttttcc ataggctcgc ccccctgac gagcatcaca aaaatcgag 7440
ctcaagtcat aggtggcgaa acccgacagg actataaaga taccaggcgt tccccctgg 7500
aagctccctc gtcgcctctc ctgttccgac cctgcgctt accggatacc tgtecgctt 7560
totccctcgg ggaagcgtgg cgtttctca atgtccacgc tgtaggtatc taagttcgtg 7620
glaggtyct cgtlccaaagc tgggctgltg gcacgaacc cccgttacc ccgacogctg 7680
cgcttatcc ggttaactatc gttctgagtc caaccggta agacagact tatcgccat 7740

```

```

ggcagcagcc actggttaaca ggattagcag agcgaggat gtagggcgtg ctacagagt 7800
cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggtg tctgcgctct 7860
gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 7920
cgctggtagc ggtgggtttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 7980
tcaaggaagat cctttgatct tttctacggg gtctgcagct cagtggaaacg aaaactcacg 8040
ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaatta 8100
aaaaagaagt tttaaatcaa tctaaagtat atatgagtaa acttggctcg acagttacca 8160
atgcttaaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcac ccatagttgc 8220
ctgactcccc gtctgttaga taactacgat acgggaaggc ttaccatctg gcccagtgcc 8280
tgcaatgata ccgcgagacc cagctccacc ggctccagat ttatcagcaa taaaccagcc 8340
agccggaagg gccgagcgca gaagtggctc tgcaacttta tccgcctcca tccagtctac 8400
taattgttgc cggaagagta gagtaagtag ttccgcaagt aatagtttgc gcaacgttgt 8460
tgccattgtc acaggcatcg tgglylccag ctgctcgttt ggtatggctt catlcagctc 8520
cgggtcccaa cgtacaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggtag 8580
ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggc 8640
tatggcagca ctgcataaatt ctcttaactgt catgccatcc gtaagatgct tttctgtgac 8700
tggtgagtag tcaaccaagt catctcgaga atagtgtatg cggcgacgga gttgctcttg 8760
cccggcgta atacgggata ataccgcgc acatagcaga actttaaaag tgcctcatcat 8820
tggaatacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc 8880
gatgtaaccc actcgtgcac ccaactgata ttccagcatc tttactttca ccagcgtttc 8940
tggtgtagca aaaaacaggaa ggcaaaaatgc cgcaaaaaag ggaataaggc cgacacggaa 9000
atgttgaata ctcatactct tcttttttca atattattga agcattttac aggggttattg 9060
tctcatgagc ggatacatat ttgaatgtat ttgaaaaat aaacaaatag ggggtccgcg 9120
cacatttccc cgaagagtag caactgacgt cgacggatcg ggagatctcc cgatccccca 9180
tggtcgactc tcagtacaat ctgctctgat gccgcatagt taagccagta tctgctccct 9240
gcttctgtgt tggaggtcgc tgagttagtc gcgagcaaaa ttaagctac aacaaggcaa 9300
ggcttgaccg acaattgcat gaagaatctg cttagggtta ggcgttttgc gctgcttcgc 9360
gatgtacggg ccagatatca ccgttgacat tgattattga ctagtatta atagtaatca 9420
attacggggt cattagtcca tagcccatat atggagttcc gcgttacata acttaacggt 9480
aatggccccc ctggctgacc gcccaacgac ccccgcccat tgacgtcaat aatgacgtat 9540
gttcccatag taacgccaat agggactttc cattgacgtc aatgggttga ctatttacgg 9600
taaaactgcc acttggcagt acatcaagtg tatcatatgc caagtacgcc cctattgac 9660
gtcaatgagc gtaaatggcc cgctggcat tatgccagat acatgacctt atgggacttt 9720
cctacttggc agtacatcta cgtattagtc atcgctatta coattggtgat cgggttttgg 9780
cagtacatca atgggcgttg atagcgtttt gactcacggg gatttccaag tctccacccc 9840
attgacgtca atgggagttt gttttggcac caaaatcaac gggactttcc aaaaatgcgt 9900
aacaactccg cccattgac gaaaatgggc ggtaggcgtg tacggtggga ggtctatata 9960
agcagagctc tctggctaac tagagaaccc actgcttact ggttatcga aattaatagc 10020
actcactata gggagaccca agcttggtag cgagctcgga tccactagta acggcccgcca 10080
gtgtgctgga attaatccg tctctgcgag ggccagctgt tgggggtgag tctccctctc 10140
aaaagcgggc atgactctcg cgctaagatt gtccagttcc aaaaacgagg aggtattgat 10200
attcactcgg cccgcggtga tgcttttag gylggccgcy tccatctggt cagaaaaagc 10260
aatctttttg ttgtcaagct tgaggtgtgg caggcttgag atctggccat ccacttgagt 10320
gacaatgaca tccactttgc ctctctctcc acaggtgtcc actcccaggt ccaactgcag 10380
gtcgatcgag ca 10392

```

<210> 18

<211> 10114

<212> DNA

<213> Artificial Sequence

<220>

<223> pESDSYNGP

<400> 18

```

tcaatatgtg ccattagcca tattattcat tgggtatata gcataaatca atattggcta 60
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gctcatgtcc 120
aatatgaccc ccattgttggc attgattatt gactagttaa taatagtaat caattacggg 180
gtcattagtt catagcccat atattggagt ccgcgttaca taacttacgg taaatggccc 240
gcctggtcga ccgcccaacg acccccggcc attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagccc cccctatttg acgtcaatga 420
cggtaaatgg cccgcctggc attatgccca gtacatgacc ttacgggact ttctacttgg 480

```

```

gcagtagatc tacgtattag tcatcgctat taccatggtg atgcggtttt ggcagtagac 540
caatgggggt ggatagcggg ttgactcacg gggattttcc agtctccacc ccattgacgt 600
caatgggggt ttgttttggc accaaaatca accgggacttt ccaaaatgtc gtaacaaactg 660
cgatcgcccg ccccgttgac gcaaaalggc ggtaggcgty tacgglggga gglctatata 720
agcagagccc gtttagtgaa ccgtcagatc actagaagct ttattgcggt agtttatcac 780
agttaaattg cttaacgagt cagtgcctct gacacaacag tctcgaactt aagctgcagt 840
gactctctia agtagcctt gcagaagtgt gtcgtgagcg actgggcagg taagtatcaa 900
ggttacaaga cagggttaag gagaccaata gaaactgggc ttgtcgagac agagaagact 960
cttgcgtttc tgataggcac ctattggtct tactgacatc cactttgcct ttctctccac 1020
agggtgccac tccagattca attacagctc ttaaggctag agtacttaat acgactccat 1080
atagggtcaga gaattccagg taagatgggc gatccctca cctgggtcaa agcctgaag 1140
aaactggaaa aagtcaccgt tcagggtagc caaaagctta ccacaggcaa ttgcaactgg 1200
gcattctccc tggtgagctt ttccacgac actaatttcg ttaaggagaa agattggcaa 1260
ctcagagagc tgatccccc ctgggagac gtgacccaaa cattgtctgg gcaggagcgc 1320
gaagcttttc agcgacctg gtgggacct agcgagctca aaatggggct gcaaatcaac 1380
aacgtgggtt agcgttaagc tagctttcaa ctgctccgcg ctaagtaoga gaagaaaaac 1440
gccacaaga acaatccga acctagcgag gagtacccaa ttatgatcga cggcgccggc 1500
aataggaaat tccgcccact gactcccagg gggtatacca cctgggtcaa caccatccag 1560
acaaagggac ttttgaagca agcctcccag aaactgttcg gcatcctgtc tgtggactgc 1620
acctccgaag aaatgaatgc ttttctcgac gtggtgccag gacaggtcgg acagaaacag 1680
atctctctgc atgcccattg caagatcgcc gacgactggg ataactcgca cccctgccac 1740
aacggccctc tgggtggctcc ccacagggg cctatcccta tgacogctag gttoattagg 1800
ggactcgggg tgcccccgga accccagatg gagccagcat ttgaccaatt taggcagacc 1860
tacagcagt ggtatcatgc agccatgagc gaggggatta aagtcatgat cggaaagccc 1920
aaggccacga acatcagcca gggggccaa gaaacctacc ctgagtttgt cgacaggtct 1980
ctgtcccaga ttaaatccga aggccaccct caggagatct ccaagttctt gacagacaca 2040
ctgactatcc aaaatccaaa tgaagagtgc agaaacgcc tgaggccact cagactcgaa 2100
gataccctgg aggaagaaat gtacgcattg cgcgacattg gcaactccaa gcaaaagatg 2160
atgctcctgc ccaaggctct gcaaacggcg ctggctggct cahtcaagg agggagactg 2220
aaggggaggt cattgaagca tgttataatt gtgggaagcc aggacattla 2280
tctagtcaat gtatagcacc taaagtctgt tttaaatgta aacagcctgg acatttctca 2340
aagcaatgca gaagtgttcc aaaaacggg agcaagggg ctcaagggag gccccagaaa 2400
caaatcttcc ngatacaaca gaagagtac cacaacaaat ctgtgtaca agagactctc 2460
cagactcaaa atctgtacc ccacatgagc gaaatacaaa tgtcaaggag 2520
aaggatcaag tagagatct caacctggac agtttgtggg agtaacatac aatctcgaga 2580
agaggcccac taccatcgct ctgatcaatg acacccctct taatgtgctg ctggacaccg 2640
gagcgcacac cagcgtttct actactgttc actataacag actgaaatad agaggaaaga 2700
aataccaggg cacaggcatc atcggcgttg gaggcaacgt cgaaaccttt tccactcctg 2760
tcaccatcaa aagaaggggg agacacatta aaaccagaat gctggctgccc gacatccccc 2820
tcaccatcct tggcagagac atttccagg acctgggcgc taaactcgtg ctggcacaac 2880
tgtctaagga aatcaagttc ngcaagatcg agotgaaaga gggcacaatg ggtccaaaaa 2940
tccccagtg gccctgacc aaagagaagc ttgagggggc taaggaaatc gtgcagcgcc 3000
tgctttctga gggcaagat agcgaggcca gcgacaataa cccttacaac agccccatct 3060
ttgtgattaa gaaaaggagc ggcaaatgga gactcctgca ggacctgagg gaaactcaac 3120
agaccgtcca ggtcggaact gagatctctc gcggactgce tcaccccgcc ggcctgatta 3180
aatgcaagca catgacagtc cttgacattg gagacgctta ttttaccatc cccctcgatc 3240
ctgaatttgc cccctatact gctttttacca tcccagcat caatcacca gaccccgata 3300
aacgtatgt gtggaagtgc ctccccagg gatltgtgct tagccctac atttaccaga 3360
agacacttca agagatctct caacctttcc gcgaaagata cccagagggt caactetacc 3420
aatatatgga cgacctgttc atggggtcca accgggtotaa gaagcagcac aaggaactca 3480
tcatcgaaat gagggcaatc ctcttgagga aaggcttcga gacacccgac gacaagctgc 3540
aagaagtccc tccatatagc tggctgggct accagctttg cctgaaaac tggaaagtcc 3600
agaagatgca gttggatatg gtcaagaacc caactctgaa cgcgtccag aagctcatgg 3660
gcaatattac ctggatgagc tccggaatcc ctgggcttac cgttaagcac attgcccga 3720
ctacaaaagg atgocctggag ttgaaccaga aggtcatttg gacagaggaa gctcagaagg 3780
aactggagga gaataatgaa agatttaaga atgtcgaag gctccaatac tacaalccc 3840
aagaagaaat gttgtgcag gtcgaaatca ctaagaacta cgaagccacc tatgtcatca 3900
aacagtcaca aggcattctg tgggccggaa agaaaatcat gaaggccaac aaaggctggt 3960
ccaccgttaa aaatctgat ctctctgctc agcacgtcgc caccgagttc atcacccgcc 4020
tcggcaagtg cccacacttc aaagtccct tcaactaagg gcaggtgatg tgggagatgc 4080
aaaaaggctg gtaactactc tggcttccc agatcgteta cccccccaa gtgggtgcag 4140
acgactggag aatgaagctt gtcgaggagc cactagcgg aattacaatc tataccgacg 4200
gcggaaaagca aaacggagag ggaatcgctg catacgtcac atctaaccgg cgcaccaagc 4260
aaaagaggct cggccctgtc actcaccagg tggctgagag gatggctatc cagatggccc 4320

```

ttgaggacac	tagagacaag	caggtgaaca	ttgtgactga	cagctactac	tgctggaaaa	4380
acatcacaga	gggctctggc	ctggaggagc	cccagctctc	ctgggtggcc	atcatccaga	4440
atatccgcga	aaaggaatct	gtctatttcc	ctcgggtgcc	tggaacacaa	ggaatttacc	4500
gcacccaact	cgccgatgaa	gcgcacaaa	ttaaagagga	aatcatgctt	gcctaccagg	4560
gcacacagat	taaggagaag	agagacgagg	acgtctggctt	tgacctgtgt	gtgccatacg	4620
acatcatgat	tcocgttagc	gacacaaaga	tcattccaac	cgatgtcaag	atccaggtag	4680
caccocaattc	atttggttgg	gtgacoggaa	agtcaccgat	ggctaagcag	ggctcttctga	4740
ttaacggggg	aatcattgat	gaaggataca	ccggcgaaat	ccaggtgtac	tgacacaaata	4800
tcggcaaaa	caatattaa	cttatcgaag	ggcagaagtt	cgctcaactc	atcatctctc	4860
agcaccacag	caattcaaga	caaccttggg	acgaaaacaa	gattagccag	agaggtgaca	4920
agggcttcgg	cagcacagg	gtgttctggg	tggaagaacat	ccagggaagc	caggacgagc	4980
acgagaattg	gcacacctcc	cctaagattt	tggcccgcaa	ttacaagatc	ccactgactg	5040
tggtcaagca	gatcacacag	gaatgcccc	actgcaccaa	acaaggttct	ggcccccgcc	5100
gctgcgtgat	gaggtcccc	aatcactggc	aggcagattg	cacccacctc	gacacacaaa	5160
ttatcttgac	cttcgtggag	agcaattccg	gctacatcca	cgcaacaactc	ctctccaaag	5220
aaaaatgcatt	gtgcacctcc	ctcgcaattc	tggaaatgggc	caggctgttc	ctctccaaat	5280
ccctgcacac	cgacaacggc	accaactttg	tggtctgaacc	tggtgtgaat	ctgctgaagt	5340
tcctgaaaa	cgccccaccc	actggcattc	cttatccacc	tgaaagccag	ggcattgtcg	5400
agagggccaa	cagaactctg	aaagaaaaa	tccaatctca	cagagacaat	acacagacat	5460
tggaaggccg	acttcagctc	gcccttatca	cttgcaacaa	aggaagagaa	agcatggggc	5520
gccagaccgc	ctgggaggct	ttcatcacta	accaggccca	ggctatccat	gaaaagctgc	5580
tccttgacga	ggcccagttc	tcacaaaagt	ctcgttttta	taagatcccc	ggtagcacg	5640
actggaaagg	tcctacacga	gttttgggga	aaggagacgg	cgagttgtg	gtgaacgatg	5700
agggcaaggg	gatcatcgct	gtgcccctga	cacgcaccaa	gcttctcatc	aagccaaact	5760
gaacccgggg	cgcccgcttc	cttttagtga	gggttaatgc	ttcgagcaga	catgatgaag	5820
tacattgatg	agtttggaca	aaccacaact	agaaatgcagt	gaaaaaatg	ctttatttgt	5880
gaaatttgg	atgctatfgc	tttatfthga	acattataa	gntgcaataa	anaagttaac	5940
aaacaaactt	gcattctatt	tatgtttcag	gttcaggggg	agatgtggga	gggttttttaa	6000
agcaagtaaa	acctctacaa	atgtggtaaa	atccgataag	gatcgatccg	ggctggcgta	6060
atagcgaaga	ggcccgaccc	gatcgccctt	cccaacagtl	gcgcagcctg	aatggcgaa	6120
ggaecggccc	tgtagcggcg	cattaaagcg	ggcgggtgtg	gtggttacgc	gcagcgtgac	6180
cgctacactt	gccagcgccc	tagcgccccc	tcctttcgtc	ttcttccctt	cccttctcgc	6240
caggttcgcc	ggctttccccc	gtcaagctct	aaatcggggg	ctccccttag	gggtccgatt	6300
tagagcttta	cgccacctcg	accgcaaaaa	acttgatttg	ggtagtggtt	cagctagtgg	6360
gcctagccgc	tgatagacgg	tttttgcgcc	tttgacgttg	gagtcacagt	tccttaaatg	6420
tggaactctg	ttccaaaactg	gaacaacact	caaccctatc	tcgggtctatt	cttttgattt	6480
ataaggaggt	ttgcgcgatt	cgccctattg	gttaaaaaat	gagctgattt	aaacaaatatt	6540
taacgcgaat	tttaacaaaa	tattaaogtt	tacaatttgc	ccgtatgcgg	tattttctcc	6600
ttacgcactc	gtgcggtatt	tcacacccga	taecgggac	tgccgagcac	catggcctga	6660
aataaactct	gaaagaggaa	cttggttagg	taccttctga	ggcggaagaa	accagctgtg	6720
gaatgtgtgt	cagttagggc	gtggaaagtc	cccaggctcc	ccagcaggca	gaagtatgca	6780
aagcatgcac	ctcaatttag	cagcaaccug	gtgtggaaag	tcceccaggc	cccacagcgg	6840
cagaagtatg	caaaagcatg	atctcaattt	gtccagcaacc	atagtcccgc	ccctaacctc	6900
gcccaccccg	ccctcaactc	cgcccagttc	cgcccattct	ccgcccattg	gctgactaat	6960
tttttttatt	tatgcagagg	ccgagggccc	ctcgccctct	gagctattcc	agaagtatgt	7020
aggagcttct	tttgagggcc	taggcttttg	caaaaagctt	gattctctcg	acacaacagt	7080
ctcgaaactt	aggtatagag	caacatgatt	gaacaagatg	gattgacagc	aggtttctccg	7140
gcgcgttggg	tggaagagct	attcggtat	gaactgggcac	aacagacaat	cggtctgctc	7200
gatgcgcgcg	tggttcggctg	ctcagcgcag	ggcgcccgcc	ttctttttgt	caagaccgac	7260
ctgcgcgggtg	ccctgaatga	actgcaggac	gagcgagcgc	ggctatcggt	gctggccacg	7320
acggggcggtc	cttgccgcagc	tggtctcgac	gttgtaactg	aagcgggaa	ggactggctg	7380
ctattggggc	aaagtgcggg	gcaggatctc	ctgtcatctc	accttgctcc	tgccgagaaa	7440
gtatccatca	tggtctgatg	aatgcggggc	ctgcatacgc	ttgatccggc	tacctgcccc	7500
ttcgaccacc	aagcgaacaa	tcgcacagag	cgagcacgta	ctcggtggga	agccggtctt	7560
gtcgatcagg	atgatctgga	cgaagagcat	caggggctcg	cgccagcaga	actgttgcgc	7620
aggctcaagg	cgccgatgcc	cgacggcgag	galctcgtcg	tgacccatgg	cgalgcctgc	7680
ttgcgaata	tcabggtgga	aaatggccgc	ttttctggat	tcactcgactg	tgcccggtcg	7740
gggttgggcg	accgctatca	ggacatagcg	ttggctaccc	gtgatattgc	tgaagagctt	7800
ggcgcggaat	gggtcgacgg	cttctcgtg	ctttacggta	tcgcgcgtcc	cgatttcgag	7860
cgcatcgctc	tcctatcgct	tccttgacag	ttcttctgag	cgggactctg	gggtttcgaa	7920
tgaccgacca	agcgagccgc	aaactgccat	cacgatggcc	gcaataaaat	atctttattt	7980
tcattacatc	tggtgtgtgg	ttttttgtgt	gaatcgatag	cgataaagat	ccgcgtatgg	8040
tgcactctca	gtacaactcg	ctctgatgcc	gcatagttaa	gccagccccc	acacccgcc	8100
acaccgcgtg	acgcgcctcg	acgggcttgt	ctgctccggg	catccgctta	cagacaagct	8160
gtgacggctc	ccgggagctg	catgtgtcag	agggttttcc	cgctcatccc	gaaacgcgcg	8220
agacgaagg	gctcgtgat	acgcctattt	ttatagggtt	atgtcatgat	aaataatggt	8280
tccttagacgt	caggtaggac	ttttcgggga	aatgtgcgcg	gaacccctat	ttgtttattt	8340
ttctaaatac	attcaaatat	gtatccgctc	atgagacaat	aacctgata	aatgcttcaa	8400
taattattga	aaaggagaag	tatgagtatt	caacatttcc	gtgtcgccct	tattcccttt	8460
tttgccgcat	tttgcttcc	tggttttgct	cacccagaaa	cgctgggtga	agtaaaagat	8520
gctgaagatc	agttgggtgc	acgagtggtg	tacatcgaac	tggaatctca	cagcggtta	8580
atccttgaga	gttttcgccc	cgaagaaagt	tttccaatga	tgagcacttt	taaaagtctg	8640
ctatgtggcg	cggtattatc	ccgtattgac	gccgggcaag	agcaactcgg	tcgcccgcata	8700
cactattctc	agaatgactt	ggtttgatgc	tcaccagtc	cagaaaaagc	tccttacggat	8760
ggcatgacag	taagagaatt	atgcagtgct	gccataacca	tgagtgataa	cactgcggcc	8820
aacttacttc	tgacaaagat	cggaggaccg	aaggagctaa	ccgctttttt	gcacaaacat	8880
ggggatcatg	taactcgctc	tgatcgttgg	gaacccggag	tgaatgaagc	cataccaaac	8940
gacgagcgtg	acaccacgat	gcctgtagca	atggcaacaa	cgttgcgcaa	actattaaat	9000
ggcgaactac	ttactctagc	ttcccgccaa	caattaatag	actggatgga	ggcggaataa	9060
gttgaggagc	cacttctgcg	ctcgccctct	ccggctggct	gggttatgtc	tgataaaatc	9120
ggagccgggtg	agcgtgggtc	tcgcggtatc	attgcagcac	tgggggccaga	tggttaagccc	9180
tcocgtatcg	tagttatcta	cacgacgggg	agtccaggca	ctatggatga	acgaaataga	9240
cagatcgctg	agataggtgc	ctcactgatt	aagcatlggt	aactgtcaga	ccaaagttac	9300
tcataataac	tttagattga	tttaaaactt	cattttttaa	ttaaaaggat	ctaggtgaag	9360
atcctttttg	ataactctat	gacccaaatc	ccttaacgtg	agttttcgtt	ccactgagcg	9420
tcagaccocg	tagaaaaagt	caaaagatct	tccttgagatc	ctttttttct	gcgcgtaatc	9480
tgctgcttgc	aaacaaaaaa	accacccgta	ccagcggtgg	ttgttttgcc	ggatcaagag	9540
ctaccaactc	tttttccgaa	ggtaactggc	ttcagcagag	cgagataacc	aaataactgt	9600
ctctagtgtg	agcgttagtt	agggccacac	ttcaagaaat	ctgtagcacc	gcctacatac	9660
ctcgctctcg	taactcctgt	accagtggtc	gctgccagt	gngataagtc	gtgtcttacc	9720
gggttggact	caagacgata	gttaccggat	aaggcgccgc	ggctcgggctg	aacggggggg	9780
tcgtgcacac	agccagctgt	ggagcgaaag	acctacaccg	aactgagata	cctacagcgt	9840
gaagctatgag	aaagcgccac	gcttcccgaa	gggagaaagg	cggacaggt	tcgggttaagc	9900
ggcagggtcg	gaacaggaga	gcgcacgagg	gagcttccag	ggggaaacgc	ctgggtatctt	9960
tatagtcctg	tcgggttttcg	ccacctctga	cttgagcgtc	gattttttgt	atgctcgtca	10020
ggggggcgga	gcctatggaa	aaaocgcagc	aacgcggcct	ttttacgggt	ccctggcctt	10080
tgctggcctt	ttgctcacat	ggctcgacag	atct			10140

<211> 10384

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY8.3G FB29(-)

<400> 19

```

agatcttgaa taataaaatg tgtgtttgtc cgaaatacgc gttttgagat ttctgtcgcc 60
gactaaatcc atgtcgccgc atagtgggtg ttatcgccga tagagatggc gatattggaa 120
aaattgatat ttgaaaatat ggcataattga aaatgtcgcc gatgtgagtt tctgtgtaac 180
tgatatcgcc atttttccaa aagtgttttt tgggcatacg cगतatctgg cगतagcgct 240
tatatcgttt acgggggatg gcगतagacg actttgggtg cttggggcgat tctgtgtgtc 300
gcaaatatcg cagtttlogat ataggtgaca gacगतatga ggctatatac cगतagaggg 360
cgacatcaag ctggcacatg gccaatgcac atcgatctat acattgaaac aatattggcc 420
attagccata ttattcattg gttatatagc ataaatcaat attggctatt ggccattgca 480
tacgttggtat ccगतtcgta atatgtacat ttatatgtgc tcatgtccaa cattacgcc 540
atgttgacat tgattattga ctagtattta atagtataca attacggggg cattagtcca 600
tagcccatat atggagttcc gcगतtacata acttacggta aatggcccgcc ctggctgacc 660
gcccacagac ccccgcccat tgacgtcaat aatgacgtat gttcccatag taacgccaat 720
agggactttc cattgacgtc aatgggtgga gtatttacgg taaactggcc acttggcagt 780
acaccaaggg tatcatattg caagtccgcc cctatttgac gtcaatgagc gtaaatggcc 840
cgctcgccat tatgcccgat acatgacctt acgggacttt cctacttggc agtacatcta 900
cgatttagtc atcgctatta ccगतgtgat cgggttttgg cagtacacca atgggcgtgg 960
atagcgggtt gactcacggg gatttccaa gttccacccc attgacgtca atgggagttt 1020
gttttggcac caaaatcaac gggactttcc aaaatgtcgt aacaactgcg atcgcccgcc 1080
ccgtgacgac aaatggggcg taggcgtgta cggtgggagg tctatataag cagagctcgt 1140
ttagtgaacc gggcactcag attctgcggg ctgagtcctt tctctgctgg gctgaaaagg 1200

```

```

cctttgtaat aaatataatt ctctactcag tccctgtctc tagtttgtct gttcgagatc 1260
ctacagttgg cgcgccgaaca gggacctgag aggggcgcag accctacctg ttgaacctgg 1320
ctgatcgtag gatccccggg acagcacagg agaacttaca gaagtctctc ggaggtgttc 1380
ctggccagaa cacaggagga caggttaagat tgggagaccc tttgacattg gagcaaggcg 1440
ctcaagaagt tagagaagg gacggtaaca ggggtctcaga aattaactac tggtaactgt 1500
aatttgggcg taagtctagt agacttattt catgatacca actttgtaaa agaaaaaggac 1560
tggcagctga gggatgtcat tccatttgcg gaagatgtaa ctcacagcgt gtcaggacaa 1620
gaagagaggg cctttgaaag aacatgggtg gcaattttct ctgtaaaagt gggcctccag 1680
attaataatg tagtagatgg aaaggcatca ttcacgctcc taagagcgaa atatgaaaaa 1740
aagactgcct ataaaaagca gtctgagccc tctgaagaat atctctagaa ctagtggatc 1800
ccccggcgct caggagtcgg gaggcacagt ggccgctttg gtcgaggcgg atccggccat 1860
tagccatatt attcattggg tatatagcat aaatcaatat tggctattgg ccattgcata 1920
cgttgtatcc atatcataat atgtacattt atattggctc atgtccaaca ttaccgccat 1980
gttgacattg attattgact agttattaat agtaactaat tacggggtca ttagttcata 2040
gcccatatat ggaagtccgc gttacataac ttacggtaaa tggcccgccct ggctgacccg 2100
ccaacgaccc ccgcgccattg acgtcaalaa tgacgtlaly lcccatagla acgccaatag 2160
ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgcccac ttggcagtag 2220
atcaagtga tcatatgcc agtacgcccc ctattgacgt caatgacggg aaatggcccg 2280
ccttgcatca tgcacagac atgaccttat gggactttcc taacttggag tacatctacg 2340
tattagtcat cgttatctac atggtgatgc ggttttggca gtacatcaat gggcgtggat 2400
acgcggttga ctacggggga ttccaaagtc tccaccccat tgacgtcaat gggagtttgt 2460
tttggcaaca aaatcaagg gactttccaa aatgtcgtaa caactccgcc ccattgacg 2520
aaatggcgcg taggcattga cgggtggagg tctatataag cagagctcgt ttagtgaacc 2580
gtcagatcgc ctggagacgc catccacgct gttttgacct ccatagaaga caccgggacc 2640
gatccagct ccgcggccccc aagcttgttg ggatccacgc gtcgccacca tgggtgagcaa 2700
ggcgaggagg ctgttccacc ggggtgggtg catcctggtc gagctggacg gcgacgtaaa 2760
cgcccaacaag ttcacgctgt ccggcgaggg cgaggcgcat gccacctacg gcaagctgac 2820
cctgaagttc atctgcacca ccggcgaagct gcccggtccc tggcccccac cctgtgacac 2880
cctgacctac gngtgcaggt gcttcagcgc ctaccccgac cacatgnagc agcagcatt 2940
cttcaagtc gccatgccg aaggctacgt ccaggagcgc accatcttct tcaaggacga 3000
cggaacctac aagaccccg ccgaggtgaa gttcgagggc gacacccctg gtaaccgcac 3060
cgagctgaag ggcctgcact tcaaggagga cggcaacatc ctggggcaca agctggagta 3120
caactcaaac agccacaacg tctatatcat ggcgcacaag cagaagaacg gcatcaagg 3180
gaacttcaag atccgcacca acatcgagga cggcagcgtg cagctgcgcg accactacca 3240
gcgaacacac ccatcgccg acggcccccgt gctgctgccc gacaaccaat accctgagac 3300
ccagtcgcgc ctgagcaaa gaccccaacga gaagcgcgat cacatggctc tgctggagtt 3360
cgtgacgcgc gccgggatca ctctcgcat ggacgagctg tacaagtaaa cggcccgcca 3420
ctctagagc gacctgcagg catgcaagct tcagctgctc gagggggggc ccggtaccca 3480
gcttttgttc cctttagtga gggttaattg ccggggaagt atttatcact aatcaagcac 3540
aagtaataca tgagaaactt ttactacagc aagcacaatc ctccaaaaaa tttgtttttt 3600
acaaaatccc tggtaaacat gattggaagg gacctaactg ggtgctgtgg aagggtgatg 3660
gtgcagtagt agttaatgat gaaggaaaag gaataattgc tgtaccatta accaggacta 3720
agttactaat aaaaccaaatt tagatttgt tgcaggaaag aagaccaaac taccattgtc 3780
agctgtgttt cctgacctca atatttgtta taaggtttga tatgaatccc agggggaaac 3840
tcaaccctta ttacccaaca gtacagaaaa tctaagtgtg aggagaacac aatgtttcaa 3900
ccttatttgt atataatga cagtgaagac agcatggcag aatcgaagga agcaagagac 3960
caagaatgaa cctgaaagaa gaatctaaag aagaaaaaaq aagaaatgac tgggtgaaaa 4020
taggtatgtt tctgttatgc ttacaggaa ctactggagg aatactttgg tggatatgaa 4080
gactccacca gcaacattat atagggttgg tggcgatagg gggaaagatta aacggatctg 4140
gccaatcaaa tgctatagaa tgctggggtt ccttcccggg gtgtagacca ttccaaaatt 4200
acttcagttt agagacatgc agaatgataa taatactgct acattattag 4260
aagctttaac caatataact gctctataaa taacaaaaaa gaattagaaa catggaagtt 4320
agtaaaagct cctggcataa ctoccttacc tatcttctct gaagctaaac cgggactaat 4380
tagacataag agagattttg gtataagtc aatagtgcca gctattgtag ccgctactgc 4440
tattgtctgt agcgtactta tgtcttatgt tgctctact gaggttaaca aaataatgga 4500
agtaaaaaat catacttttg aggtagaaaa tagtactcta aatggtatgg atttaataga 4560
acqacaaata aagatattat atgctatgat tcttcaaaac catgcagatg ttcaacttgt 4620
aaaggaagga caacaggtag aggagacatt taatttaatt ggatgtatag aaagacacaa 4680
tgtattttgt catactggtc atccctggaa tatgtcatgg ggaacattta atagctcaac 4740
acaatgggat gactgggtta gcaaaatgga agatttaaat caagagatac taactacact 4800
tcatggagcc aggaacaatt tggcacacac catgataaca ttoaatacac cagatagtat 4860
agctcaattt ggaaagacc tttggagtca tattggaaat tggattctcg gattgggagc 4920
ttccattata aaatatatag tgatgttttt gcttatttat ttgttactaa cctcttcgcc 4980
taagatcttc agggccctct ggaaggtgac cagtgggtga gggctctccg gcagtctgta 5040

```

```

cctgaagaaa aaattccatc acaaacatgc atcgcgagaa gacacctggg accagggcca 5100
acacaacata cacctagcag cgtgacccgg tggatcaggg gacaaatact acaagcagaa 5160
gtactccagg aacgactgga atgggagatc agaggagtag aacaggccggc caaagagctg 5220
ggtgaagtca atcgaggcat ttggagagag ctatatctcc gagaagacca aaggggagat 5280
ttctcagcct gggggcgcta tcaacgagca caagaacggc tctgggggga acaatcctca 5340
ccaagggtcc ttagacctgg agattcgaag cgaaggagga aacatttatg actgttgcat 5400
taaaagccaa gaaggaaatc tcgtatccc ttgctgtgga ttcccttat ggctattttg 5460
gggactagta attatatag gacgcatagc aggtatgga ttactgtgac tcgtgttat 5520
aataaggatt tgtattagag gcttaaatit gatatttgaa ataacagaa aaatgcttga 5580
ttatattgga agagctttaa atcctggcac atctcatgta tcaatgcctc agtatgttta 5640
gaaaaacaag gggggaactg tgggtttttt atgaggggtt ttataaatga ttataagagt 5700
aaaaagaaag ttgctgatgc tctcataacc ttgtataacc caaaggacta gctcatgttg 5760
ctaggccaact aaacgcgaat aacgcgattt gtgacgcgag ttcccatgt gtgacgcgtt 5820
aacttctgt ttttacagta tataagtgt tgtattctga caattgggca ctgagattct 5880
gcggtctgag tccctctct gctgggctga aaaggccttt gtaataata taattctota 5940
ctcagtcctt gctcttagtt ttgtgttct agatcctaca gagctcatgc cttggcgtaa 6000
tcatgttcat agtcttctc tgttctcgc tcttatccgc tcacaattcc acacaacata 6060
cgagccggaa gctaaagtgt taaagcctgg ggtgcctaat gactgagcta actcacatta 6120
attgccttgc gctcactgct cgtttccag tcgggaaacc tgtctgcca gtgatgccc 6180
ggcgcccgag gcggcctacg tgaacctaca cccaaatcaa gttttttgct gtcgaggtgc 6240
cgtaaagctc taaatcgaaa cctaaagggt agcccccgat tttagagctt acgggggaaa 6300
ccggcgaaag tggcgagaaa ggaagggaag aaagcgaaaag gagcgggcgc tagggcgctg 6360
gcaagtgtag cgttcacgct gcgcgttaacc accacacccg ccgcgcttaa tgcgcgcta 6420
cagggcgcgt ccattcgcca ttccggctgc gcaactgttg ggaaggcgca tgggtgcggg 6480
cctctctgct attacgcag cccggatcga tcttatctgg attttaccac atttgtagag 6540
gtttactctg ctttaaaaaa cctccacat ctccccctga acctgaaaca taaatgaat 6600
gnaattgttg ttgttaactt gtttaattga gttataatg gttacaaata aagcaataga 6660
atocaaatc toacaaatc agcatttttt tcactgcatt ctagtgtgtg ttgttccaaa 6720
ctcatcaatt tatcttctca tgtctgctgc aagcattaac cttcaataaa gggaagcggc 6780
cgccccggtc gacttcaacg gtgtttgctg cgttttttgg agtctccggg cctcaagacg 6840
cggggctgct tctgctcgc acacagcctt tctgtgccc tctggtagcc tcccatagc 6900
gagaatctgc cctctagctc ctccgggaag tagagctccc tccagatgcc gcgattcacc 6960
tctccagctc ctttagcggc ttgttgcaag cccctaattc tccattccag cettctctgg 7020
aggacctcgg cttgcaaaat ctggccccta atccacctat ccttctgga ggggtgtgtc 7080
tgggtgggac gggggcgag gtgtctctct gctgtgcaag tctggctagg aatcttctcc 7140
tcgggcaggg actgtctcag cagcgccac cactggctcc cctccagggg gcttctgtgg 7200
tcgatcttcc accagctgtt gcggcgcttc tctcttttgc tctctcctt gaggttcata 7260
tcttgatccc tggcctcctt gctctcagcc atgtggcgga attctcgagg ctgacctccc 7320
ggtgtgtgtt ggtgtgtccc tggcgagggg tctccagatc ccggacgagc ccccaaatga 7380
aagacccccc agacgggtag tcaactcctc tgaggagacc ctcccaagga acagcgagac 7440
cacgagctcg atgcaacagc aagaggattt attgataca cgggtaccgc ggcgactcag 7500
tctatcgag gactggcgcc ccgagttagg ggttgtgagc tctttatag agctcgggaa 7560
gcagaaagcg gcgaaacagaa gcgagaaagca ggtctgattg ttaattcaaa taaggcacag 7620
ggctatttca ggtccttggg ggagcctgga aacatctgat ggtgtctaa aaactgctga 7680
gggtgtgggc atatctgggg accatctgtt cttggcccgc ggcgggggac gaacggcggt 7740
gacctctgt tcttgcccc cggccggggc cgaaactgct caccgcagat atcctgtttg 7800
gcccacgctt agctgttttc gtgtaacgc ctttgtatctg aacttctcta tcttctgttt 7860
ggatatttcc catgctctgc aaaatggcgt taactggcgt atcaggctaa gcaatttgag 7920
atctggcgga ggcggcctac tctgcattaa tgaatcgccc aacggcgggg gagaggcggt 7980
ttgcgtattg ggcgctcttc cgttctctgc ctcaactgact cgtcgctcgc ggtcgttcgg 8040
ctgcggcgag cgttatcagc tcaactcaag gcggtaatac ggttatccac agaatcaggg 8100
gataacgcag gaagaaacat gtataacttc gtataatgta tgcatacga agttatacat 8160
gtgagcaaaa ggcacgcaaa agggcaggaa ccgtaaaaag gccggttgc tggcggtttt 8220
ccatagggct cgcgccctct agcgatctca caaaaatoga cgtcgaagtc agaggtggcg 8280
aaacccgaca ggaataataa gataccaggg gtttccccct ggaagctccc tctgtgctcc 8340
tctgttccgc accctgcgcg ctctcagcta cctgacgctc ttctccctc cgggaagcgt 8400
ggcgcttctc catagctcac gctgtaggta tctcagttcg gtgtaggctg tctgctccaa 8460
gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgcttat ccgtaacta 8520
tcgtcttaga tccaaacccg taagacagga ctatcgcca ctggcagcag ccaactgtaa 8580
caggatttag agagcgagg atgtaggcgg tctacagag ttcttgaagt ggtggcctaa 8640
ctacggctac actagaaggc cagtatttgg tatctgcgt ctgctgaagc cagttaacct 8700
cggaaaaaag gttggttagc cttgatccgg caaaacaaac accgctggtg cgggtgtgtt 8760
ttttgttccg aagcagcaga ttacgcgcag aaaaaaagga tctcaagaa atcttttgat 8820
ctttcttagc ggttctgacg ctcdgtggaa cgaaaaacta cgttaaggga ttttggctat 8880

```

```

gagattatca aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc 8940
aatctaaagt atatatgagt aaacttggct tgacagttac caatgcttaa tcaagtaggc 9000
acctatctca gcgatctgtc fatttctgtc atccatagtt gctgactccc cgtcgtgtga 9060
gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga tacccgcaga 9120
cccacgctca ccggctccag atttatcagc aataaaacag ccagccggaa gggccgagcg 9180
cagaagtgtt cctgcaactt tatccgcctc catccagtct attaatgttt gccgggaagc 9240
tagagttaag agtccgcag ttaatagttt ggcgaacgtt gttgccattg ctacaggcat 9300
cgtgtgtgtc cgtcgtcgt ttggtatggc ttcattcagc tccgggtccc aacgatcaag 9360
gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat 9420
cgttgtcaga agtaagtgg ccgcagttgt atcactcatg gttatggcag cactgcataa 9480
ttctcttact gtcagccat ccgtaagatg ctlttctgtg actggtgagt actcaaccaa 9540
gtcattctga gaatagtgtg tgcggcgacc gagttgtctc tgcccggcgt caatacggga 9600
taataccgcg ccacatagca gaactttaaa agtgcctatc attggaaaac gttcttcggg 9660
gcgaaaaact tcaaggatct taccgctgtt gagatccagt tccatgtaac ccaactcgtc 9720
acccaactga ctttagcatc cttttacttt caccagcgtt tctgggtgag caaaaaacag 9780
aaggcaaaat gccgcaaaaa agggaaabaag ggcgacacgg aaatgttgaa tactcatact 9840
cttccttttt caatatattt gaagcattta tcaagggttat tgtctcatga gccgatacat 9900
atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc ccgaaaaagt 9960
gccacctaaa ttgtaagcgt taatattttg ttaaaattcg cgttaaatit ttgttaaatc 10020
agctcatttt ttaaccaata ggccgaaaatc cttataaatc aaaagaatag 10080
accgagatag ggttgagtgt tgttccagtt tggaaacaaga gtccactact aaagaaacgt 10140
gactccaaag actagaaggc aaaaaccgtc tatcaggggc atggcccact acgtgataac 10200
ttcgtataat gtatgctata cgaagtatat actacgtgaa ccatcacctt aatcaagttt 10260
tttggggctg aggtgcctga aagcactaaa tccgaaacct aaagggagcc ccgatttag 10320
agcttgacgg ggaagccaa cctgggttat cgaaattaat acgactcact ataggagac 10380
cggc 10384

```

<210> 20

<211> 21

<212> DNA

<213> Artificial Sequence

<220>
 <223> Sequence flanking the codon-optimised EIAV gag/pol ORF

<400> 20
 tctagagaat tcgccacat g 21

<210> 21
 <211> 18
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Sequence flanking the codon-optimised EIAV gag/pol ORF

<400> 21
 tgaacccggg gcggccgc 18

<210> 22
 <211> 11
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> pEV53B

<400> 22
 caggtaagat g 11

<210> 23
 <211> 42
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<400> 23
 ggctagagaa ttccaggtaa gatgggcat ccctcacct gg 42

<210> 24
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<400> 24
 ttgggtactc ctgctaggt tc 22

<210> 25
 <211> 4307
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Codon optimised gag-pol sequence (pSYNGP)

<400> 25

```

atgggagcgcc ggcgcagcgt gctgtcgggc ggcgagctgg accgctggga gaagatccgc 60
ctgggccccg ggcgcacaaa gaagtacaag ctgaagcaca tegtgtgggc cagccgcgaa 120
ctggagcgct tgcgcgtgaa ccccgggctc ctggagacaa cgcgaggggtg cgcgcgagtc 180
ctgggccaac tgcagccccg cctgcaaac cgcagcgagg agctgcgcag cctgtacaac 240
accgtggcca cgtgtactg cgtccaccag cgcacgaaa tcaaggatac gaaagaggcc 300
ctggataaaa tcgaagagga acagaataag agcaaaaaga agccccaaca ggcgcgcggc 360
gacaccggac acagcaacca ggtcagccag aactacccca tegtgcagaa catccagggg 420
cagatgggtc accaggccat ctccccccgc acgctgaacg cctgggtgaa ggtggtggaa 480
gagaaggctt ttagnccgga ggtgataccc atgttctcag cctgtcaga gggagccacc 540
ccccagatc tgaaccccat gctcaacaca gtgggggggc accaggccgc catgcagatg 600
ctgaaggaga ccatcaatga ggaaggctgc gaaltgggac gtgtgcaacc ggtgcacgca 660
gggcccctcg caccgggcca gatgcgtgag ccacggggct cagacatcgc cggaaacgat 720
agtaccotto aggaacagat cggctggatg accaacaacc caccatccc ggtgggagaa 780
atctacaaa cgtggatcat cctgggcctg aacaagatcg tgcgcagtta tagccctacc 840
agcatcctgg acatccgcca aggcgcgaag gaaccttct gcgactacgt ggaccgggtc 900
tacaaaaacg tccgcgcgca gcaggctagc caggaggtga agaactggat gaccgaaacc 960
ctgctgggtc agaacycgaa cccggactgc aagacgatcc tgaaggccct gggccacgag 1020
gtcaccctag aggaatgat gaccgcctgt caggaggtgg gcggaccccg ccacaaggca 1080
cygctcctgg ctgaggccat gagccaggtg accaactcgc claccatcat gatgcagcgc 1140
ggcaacttct ggaaccaacg caagatcgtc aagtgcctca actgtggcaa agaaggcgac 1200
acagcccgca actgcagggc cctagggaaa aagggtctgt ggaatgctgg aaaggaagga 1260
caccaaatga aagattgtac tgagagacag gctaattttt tagggaaagt ctggccttcc 1320
cacaagggaa ggcaggggaa ttttctctag agcagaccag agccaacagc cccaccagaa 1380
gagagcttca ggtttgggga agagacaaca actccctctc agaagcagga gccgatagac 1440
aaggaaactg atcctttagc ttccctcaga tcaactcttg gcagcgaccc ctgctcaca 1500
taaaataggg gggcgagctc aaggaggtct tctggacac cggagcagac gacaccgtgc 1560
tggaaggagat gctgttgcca ggcgcctgga agccgaagat gatcggggga atcggcggtt 1620
tcatcaaggt gcgccaagt gaccagatcc tcatcgaaat ctgcggccac aaggctatcg 1680
gtaccgtgct ggtgggcctc acaccgtca acatcatcgg acgcaacctg ttgaccgaga 1740
tcggttgca cgtgaacttc cccattagcc ctatcgagac ggtaccggtg aagctgaagc 1800
ccgggatgga cggcccgaa gtcaagcaat ggccattgac agaggagaag atcaaggcac 1860

```

```

tggtggagat ttgcacagag atggaaaagg aagggaataa ctccaagatt gggcctqaga 1920
accgctacaa cagcccggtg ttgcgaatca agaagaagga ctgcagcaaa tggcgcaagc 1980
tggtggactt ccgcgagctg cgcgaagctc aacaagcgca cgcgaagact ctgggaggtt cagctgggca 2040
tcccgaccac ccgaggggtg aagaagaaga aatccgtgac cgtactggat gtgggtgatg 2100
cctaacttcc cgttccctcg gacgaagact tcaggaagta cactgccttc acaatccctt 2160
cgatcaacaa cagagacacg gggattcgat atcagtacaa cgtgctgccc cagggtctga 2220
aaggctctcc cgcgaacttc cagagtagca tgacaaaaat cctggagcct ttcgcaaac 2280
agaaccccca catcgtcatc tatcagtaca tggatgactt gtacgtgggc tctgatctag 2340
agatagggga ggcacgcacc aagatcgagg agctgcgcca gcacctgttg aggtggggac 2400
tgaccacacc cgacaagaag caccagaagg agcctccctt cctctggatg ggttacgagc 2460
tgacccctga caaatggacc gtgcagccta tegtgtctgc agagaagac agctggactg 2520
tcaacgacat acagaagctg gtggggaagt tgaactgggc cagtcaagat taccagggga 2580
ttaaggtgag gcagctgtgc aaactcctcc gcggaaccaa ggcaactaca gaggatgatc 2640
ccctaaccga ggaaggccag ctcgaaactg cagaaaaccg agagatccta aaggagcccg 2700
tgacaggcgt gtactatgac cctcccaagg acctgatcgc cagatccag aagcaggggc 2760
aaggccagtg gacctatcag atttaccagg agcccttcaa gaacctgaag accggcaagt 2820
acgcccggat gaggggtgcc cacactaacg acgtcaagca gctgacagag gccgtgcaga 2880
agatcacacc cgaaagcacc gtgatctggg gaaagactcc taagttcaag ctgccatccc 2940
agaaggaaac ctgggaaacc tggtggacag agtattggca ggccacctgg attcctgagt 3000
gggagttcgt caacacccct cccctgggtg agctgtggtg ccagctggag aaggagccca 3060
tagtgggcgc gaaaaccttc tacgtggatg gggccgctaa caggagagact aagctgggca 3120
aagccggata cgtcactaac cggggcagac agaaggttgt caacctcact gacaccacca 3180
accagaagac tgagctgcag gccatttacc tgcctttgca ggactcgggc ctggaggtga 3240
acatcgtgac agactctcag tatgcctagg gcatcatcaa agcccagcca gaccagagt 3300
agtcaggact ggtcaatcag atcatcgag agctgatcaa gaaggaagaa gtcctatctg 3360
cctgggtacc cgcaccacaa ggcatggcg gcaatgagca ggtcgacaag ctggtctcgg 3420
ctggcatcag gaaggtgcta ttctggatg gcatcgacaa ggcccaggac gagcacgaga 3480
aataccacag caactggcgg gccatggcta ggcacttcaa cctgcccctt gtggtggcca 3540
aagagatcgt ggcagctgt gacaagtgc agctcaaggg cgaagccatg catggccagg 3600
tggaactgtg ccccgcatc tggcaactcg attgcacca tctggagggc aaggttatcc 3660
tggtagccgt ccatgtggcc agtggctaca tcgaggccga ggtcatctcc gccgaacag 3720
ggcaggagac agcctacttc ctctgaagc tggcagggcg gtggccagtg aagaccatcc 3780
atactgacaa tggcagcaat ttaccagtg ctacgggtta ggccgcctgc tgggtggcgg 3840
gaatcaagca ggaattcggg atcccctaca atcccagag tcagggctgc gtcgagtcta 3900
tgaataagga gttaaagaag attatcgccc aggtcagaga tcaggctgag catctcaaga 3960
ccgcggtcca aatggcggtg ttcatccaca atttcaagc gaaggggggg attgggggt 4020
acagtgcggg ggagcggatc gtggacatca tcgcgaccga catccagact aaggagctgc 4080
aaaagcagat tacaagatt cagaatttcc ggtctacta cagggaacgc agaaaatccc 4140
tctgaaagg cccagcgaa ctctctgga aggtgaggg ggcagtagtg atccaggata 4200
atagcgacat caaggtgtg cccagaagaa aggcgaagat cattagggat tatggcaaac 4260
agatggcggg tgatgattgc gtggcgagca gacaggatga ggattag 4307

```

<210> 26

<211> 4307

<212> DNA

<213> Artificial Sequence

<220>

<223> pGP-RRE3

<400> 26

```

atgggtgcga gaggcgtcagt attaagcggg ggagaattag atcgatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataaa ttaaaacata tagtatgggc aagcaggggag 120
ctagaacgat tccgagttaa tccctggcctg ttagaacatc cagaaggctg tagacaaaata 180
ctgggacagc tacaaacctc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtgcgaa cctcttatgt tgtgcataaa aggatagaga taaaagacac caaggaagct 300
ttagacaaga tagaggaaga gcaaaacaaa agtaagaaaa aagcacagca agcagcagct 360
gacacaggac acagcaatca ggtcagccaa aattacccta tagtgacagaa catccagggg 420
caaatggtag atcaggccat atcacctaga acttttaaatg catgggtaaa agtagtagaa 480
gagagggtct tcagcccgaga agtgataccc atgttttcag cattatcaga aggagccacc 540
ccacaagatt taaacacctg gctaaacaca gtgggggggac atcaagcagc catgcaaatg 600
ttaaaagaga ccatcaatga ggaagctgca gaattgggata gagtgcaccc agtgcacgca 660
gggcctattg caccaggcca gatgagagaa ccaaggggaa gtgacatagc aggaactact 720

agtaggagaa 780
atttataaaa gatggataat cctgggatta aataaaatag taagaatgta tagccctacc 840
agcattctgg acataagaca aggacaaaaa gaacccttta gagactatgt agaccgggtc 900
tataaaactc taagagccga gcaagcttca caggaggtaa aaaattggat gacagaaacc 960
ttgttggtcc aaaaatgcgaa cccagattgt aagactattt taaaagcatt gggaccagcg 1020
gtcactactg aagaatgtat gacagcatgt cagggagtag gaggaccccg ccataaggca 1080
agagttttgg ctgaagcaat gagccaaagta acaaatccag ctaccataat gatgcagaga 1140
ggcaatttta ggaaccaaag aaagatttgt aagtgtttca attgtggcaa agaagggcac 1200
acagccagaa attgcagggc cctagggaaa agggctgttt ggaaatgtgg aaaggaagga 1260
caccaaatga aagattgtac tgagagacag gctaattttt taggggaagt ctggccttcc 1320
tacaaaggaa ggccagggaa ttttcttcag agcagaccag agccaacagc cccaccagaa 1380
gagagcttca ggtctggggt agagacaaca actcccccctc agaagcagga gccgtagagc 1440
aaggaactgt atcctttaac tccccccaga tccctcttgg gcaacgaccc ctctgcacaa 1500
taaaagtagg ggggcaacta aaggaagctc taltagatac aggcagcagat gatcagtat 1560
tagaagaatg gagtttgcca ggaagatgga aaccaaataat gataggggga attggagggt 1620
ttatcaaatg aagacagtat gatcagatac tcatagaatct ctgtggacat aaagctatag 1680
gtacagtatc agtaggacct acacctgtca acataattgg aagaatactg ttgactcaga 1740
ttggttgcac tttaaatttt cccattagcc ctattgagac tttaccagta aaattaaagc 1800
caggaaatga tggcccaaaa gttaaaacat ggccattgac agaagaaaaa ataaaagcat 1860
tagtagaata ttgtacagag atggaaaagg aagggaaaaa ttcaaaaatt gggcctgaaa 1920
atccatcaca tactccagta ttgtccataa agaaaaaaga cagtlactaaa tggagaaaaa 1980
tagtagattt cagagaactt aataagagaa ctcaagactt ctggggaagt caattaggaa 2040
taccacatcc cgcagggtta aaaaagaaaa aatcagtaac agtactggat gtgggtgatg 2100
catatttttc agttccctta gatgaagact tcaggaaata tactgcattt accataccta 2160
gtataaacaa ttgagacacca gggattagat atcagtacaa tgtgcttcca cagggaatga 2220
aaggatcacc agcaatatcc caaagtagca tgacaaaaat cttagagcct tttagaaaac 2280
aaaaatccga catagttatc tatcaataca tggatgattt gtatgtagga tctgacttag 2340
aaatagggca gcatagaaca aaaaatagagg agctgagaca acatctgtgt aggtggggac 2400
ttaccacacc agacaaaaaa catcagaaag aacctccatt cctttggatg ggttatgaac 2460
tccatcctga taaatggcca gtacagccca tagtgctgcc agaaaaagac agctggactg 2520
tcaatgacat acagaagtfa gtggggaat tgaattgggc aagtgcagatt taccaggga 2580
ttaagtagag gcaattatgt aaactcctta gaggaaccaa agcactaaca gaagtaatac 2640
cactaacaga agaagcagag ctagaactgg cagaaaaacg agagatttota aaagaaccag 2700
tacatggagt gtattatgac ccatcaaaag acttaatagc agaaatacag aagcaggggc 2760
aaggccoaatg gacatatcaa atttatcaag agccatttaa aaatctgaaa acaggaaaaa 2820
atgcaagaat gaggggtgcc cacactaatg atgtaaaaa ataacagag gcagtgcaca 2880
aaataacccc agaaagcata gtaatatggg gaaagactcc taattttaaa ctgccatac 2940
aaaaggaaac atgggaaaca ttggtggacag agtattggca agccacctgg attcctgagt 3000
gggagtttgt taataccctc cctttagtga aattatggta ccagttagag aaagaacca 3060
tagtagggag agaaaacctc tatgtagatg gggcagctaa caggggagact aaattaggaa 3120
aagcaggata tgttaclaat agaggaagac aaaaagttlyl caccctlacl gacacaacaa 3180
atcagaagac tgaattacaa gcaatttato tagctttgca ggatttcggga ttagaagtaa 3240
acatagtaac agactcaaaa latgcattag gaatcattca agcacaaaca gatcaagtg 3300
aatcagagtt agtcaatcaa ataataagag agttaataaa aaaggaaaag gtctatctgg 3360
catgggtacc agcacacaaa ggaattggag gaaatgaaca agtagataaa ttagttagtg 3420
ctggaatcag gaaagtacta tttttagatg gaatagataa ggcccaagat gaacatgaga 3480
aatatcacag taattgggga gcaatggcta gtgattttaa cctgccacct gtagtagcaa 3540
aagaatatgt agccagctgt gataaatgtc agctaaaagg agaagccatg catggacaag 3600
tagactgtag tccaggaata tggcaactag attgtacaca tttagaagga aaagttatcc 3660
tggtagcagt tcatgtagcc agtggatata tagaagcaga agttattcca gcagaaacag 3720
ggcaggaaac agcatatttt cttttaaata tagcagggaag atggccagta aaaacaatac 3780
atacagacaa tggcagcaat ttaccagtg ctacgggttaa ggcgcgctgt tggtggggcg 3840
gaatcaagca ggaatttgga attccctaca atcccaaaag tcaaggagta gtagaatcta 3900
tgaataaaga attaaagaaa attataggac aggtaaagaa tcaggctgaa catcttaaga 3960
cagcagtcaca aatggcagta ttcatccaca attttaaaag aaaagggggg attggggggg 4020
acagtgcagg ggaagaata gttagacata tagcaacaga catacaaact aaagaattac 4080
aaaaacaaat tacaaaaatt caaaatttcc gggtttatta caggcacagc agaaatccac 4140
tttggaaagg accagcaaaag ctccctctgga aaggtgaagg ggcagttaga atacaagata 4200
atagtacatg aaaagttagt ccaagaagaa aagcaagat cattagggat tatggaaaaa 4260
agatggcagg tgaatgtgt gtggcaagta gacaggatga ggaattag 4307

```

<210> 27

<211> 4658

<212> DNA

<213> Equine infectious anemia virus

<400> 27

```

atggggagacc ctttgacatg gagcaaggcg ctcaagaagt tagagaaggt gacgggtacaa 60
gggtctccaga aatttaactac tggtaactgt aattgggcgc taagtctagt agacttattt 120
catgatatacca actttgttaaa agaaaaggac tggcagctga gggatgtcat tccattgtctg 180
gaagatgttaa ctccagcgtc gtccaggacaa gaaagagagg cctttgaaag aacatggttg 240
gcaattttctg ctgtaaagat gggcctccag attaataatg tagtagatgg aaaggcatca 300
ttccagctcc taagagcgaa atatgaaaag aagactgcta ataaaaagca gtctgagccc 360
tctgaagaat atccaatcat gatagatggg gctggaacaa gaaattttag acctctaaca 420
ccttagaggat atactacttg ggtgaatacc atacagacaa atggtctatt aaatgaagct 480
agtcaaaact tatttgggat attatcagta gactgtactt ctgaagaaat gaatgcattt 540
ttggatgtgg tacctggcca ggcaggacaa aagcagatat tacttgatgc aattgataag 600
atagcagatg attgggataa tagacatcca ttaccgaatg ctccactggg ggcaccacca 660
caagggccta ttcccattgc agcaagggtt attagagggt taggagtacc tagagaaga 720
cagatggagc ctgcttttga tcagtttagg cagacatata gacaatggat aatagaagcc 780
atgtcagaag gcatcaaatg gatgattgga aaacctaaag ctcaaaaat taggcaagga 840
gctaaggaa cttaccocga attttagac agactattat cccaaataaa aagtgaagga 900
catccacaag agatttcaaa attcttgact gatacactga ctattcagaa cgcaaatgag 960
gaatgtagaa atgctatgag acatttaaga ccagaggata cattagaaga gaaaatgtat 1020
gcttgagag acattgggaa tacaaaaaa aagatgatgt tattggcaaa agcacttcag 1080
actgtctctg cggggccatt taaagtgga gccttgaaag gaggggccact aaaggcagca 1140
caaatatgtt ataactgttg gaagccagga catttatcta gtcaatgtag agcacttaaa 1200
gtcgttttta aatgtataca gcctggacat ttctcaaaag aatgcagaag tgttccaaaa 1260
aacgggaagc aaggggctta agggaggccc cagaacaaaa ctttcccgat acaacagaag 1320
agtcagcaca acaaatctgt tgtacaagag actcctcaga ctcaaaatct gtaccagat 1380
ctgagcgaaa taaaaaagga aaggagaagg atcaagttaga ggaatctaac 1440
ctggacagtt tgtgggagta acatataatc tagagaaaag gcctactaca atagtattaa 1500
ttaatgatac tcccttaaat gtactgttag acacaggagc agatacttca gtgttgacta 1560
ctgcacatta taatagggtt aatatagag ggagaaaata tcaagggacg ggaataatag 1620
gagtgaggag aatgtgggaa acattttcta cgcctgtgac tataaagaaa aagggttagac 1680
acataagac aagaatgtca gtggcagata ttccagtgc tattttggga cgagatattc 1740
ttcaggactt aggtgcaaaa ttggttttgg cacagctctc caaggaaata aaatttagaa 1800
aaatagagtt aaaaagggtt acaatggggc caaaaattcc tcaatggcca ctactaagg 1860
agaaactaga agggggccaaa gagatagtc aaagactatt gtcagaggga aaaatatcag 1920
aagctagtga caataatcct tataattcac ccatatttgt aataaaaaag aggtctggca 1980
aatggaggtt attacaagat ctgagagaat taacaaaaac agtacaagta ggaacggaaa 2040
tatccagagg attgcctcac ccggggagat taattaaatg taacacatg actgtattag 2100
atattggaga tgcattttc actataccct tagatccaga gtttagacca tatacagctt 2160
tcactattcc cctccattat catcaagaac cagataaaaag atatgtgtgg aaatgtttac 2220
cacaggattt cgtgttgagc ccatatata atcagaaaac attacaggaa attttacaac 2280
cttttaggga aagatatcct gaegtacaat tgtatcaata tatggatgat ttgttcatgg 2340
gaagttaatg ttctaaaaaa caacacaaag agttaatcat agaattagg ccgatcttac 2400
tggaagagg ttttgagaca ccagatgata aattacaaga agtgccacct tatagtctggc 2460
taggttatca actttgtcct gaaaatttga aagtacaaaa aatgcaatba gacatggtaa 2520
agaatccaac ccttaatgat gtgcaaaaat taatggggaa tataacatgg atgagctcag 2580
ggatcccagg gttgacagta aaacacattg cagctactac taagggatgt ttgaggttga 2640
atcaaaaagt aatttggcgc gaagaggcac aaaaagagtt agaagaaaat aatgagaaga 2700
ttaaaaatgc tcaagggtta caattattata atccagaaga agaaatgtta tgtgaggttg 2760
aaattacaaa aaatttatgag gcaacttatg ttataaaaa atccaaagga atcctatggg 2820
caggtaaaaa gattatgaag gctaataagg gatggtcaac agtaaaaaat ttaatgttat 2880
tgttgcaaca tgtggcaaca gaaagtatta ctagagtagg aaaaatgtcca acgtttaagg 2940
taccatttca caaagagcaa gtaatgtggg aaatgcaaaa aggatgggat tattcttggc 3000
toccagaaat agtatataca catcaagtag ttcatgatga ttggagaatg aaattggtag 3060
aagaacctac atcaggaata acaatataca ctgatggggg aaaacaaaaa ggagaaggaa 3120
tagcagctta tgtgaccagt aatgggagaa ctaaacagaa aaggttagga cctgtcactc 3180
atcaagtggc tgaaggaatg gcaatacaaa tggcattaga ggataccaga gataacaag 3240
taaatatagt aactgatagt tattatttgt ggaaaaatat tacagaagga ttaggtttag 3300
aaggaccaca aagtccttgg tggcctataa tacaaaatat acgagaaaaa gagatagttt 3360
attttgcctg ggtacctggg cacaagggga tatatggtaa tcaattggca gatgaagccg 3420
caaaaaataa agaagaaatc atgctagcat accaaggcac acaaattaaa gagaaaagag 3480
atgaagatgc agggtttgac ttatgtgttc cttatgacat catgatacct gtatctgaca 3540
caaaaatcat acccacagat gtaaaaaatc aagttcctcc taatagcttt ggatgggtca 3600

```

```

ctgggaatc atcaatggca aaacaggggt tattaattaa tggaggaata attgatgaag 3660
gatatacagg agaaatacaa gtgatatgta ctaatatggg aaaaagtaat attaaattaa 3720
tagagggaca aaaattlga caatltaatta tactacagca lcaactcaaat lccagacagc 3780
cttgggatga aaataaaaata tctcagagag gggataaagg atttggaaat acaggagtat 3840
tctgggtaga aaatattcag gaagcacaag atgaacatga gaattggcat acatcaccaa 3900
agatattggc aagaaattat aagataccat tgactgtagc aaaacagata actcaagaat 3960
gtcctcattg cactaagcaa ggcacaggac ctgcagggtt tgtcatgaga tctcctaact 4020
attggcaggc agattgcaca catttgggaca ataagataat attgactctt gttagatcaa 4080
attcaggata catacatgct acattattgt caaaagaaaa tgcattatgt acttcattgg 4140
ctatttttag atgggcaaga ttgttttcc caaagtcctt acacacagat aacggcacta 4200
attttgggc agaaccagtt gtaaatttgt tgaagttcct aaagatagca cataccaag 4260
gaataccata tcatccagaa agtcaggqta ttqtaaaag qgcataatag accttgaag 4320
agaagattca aagtcataga gacacactc aaacactgga ggcagcttta caactgtctc 4380
tcattacttg taacaaaggg agggaaagta tgggagggaca gacaccatgg gaagtattca 4440
tcactaatca agcacaagta atacatgaga aacttttact acagcaagca caatcctcca 4500
aaaaatttgg tttttacaaa atccctgggt aacatgattg gaagggaact actagggtgc 4560
tgttgaaggg tgatgggtgca gtatgtagta atgatgaagg aaagggaata attgctgtac 4620
catlaaccag gactaagtta ctaataaaac caaattga 4658

```

<210> 28

<211> 385

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY3.1

<400> 28

```

atgggagacc ctttgacatg gagcaaggcg ctcaagaagt tagagaaggt gacgggtacaa 60
gggtctcaga aattaactac tggtaactgt aattgggccc taagtctagt agacttattt 120
catgatacca actttgttaa agaaaaggac tggcagctga gggatgtcat tccattgctg 180
gaagatgtaa ctcagacgct gtcaggacaa gaaagagagg cctttgaaag aacatgggtg 240
gcaatttctg ctgtaaatg gggcctccag attaataatg tagtagatgg aaaggcatca 300
ttccagctcc taagagcgaa atatgaaaag aagactgcta ataaaaagca gtctgagccc 360
tctgaagaat atccaatcat gatag 385

```

<210> 29

<211> 385

<212> DNA

<213> Artificial Sequence

<220>

<223> pONY3.2opti

<400> 29

```

atgggcgac ccctcacctg gtccaaagcc ctgaaaaaac tggaaaaagt caccgttcag 60
ggtagccaaa agcctaccac aggcacattgc aactggggcat tglccctygt ggalcttttc 120
cagcactacta atttcgttaa ggagaaagat tggcaactca gagacgtgat cccctctctg 180
gaggacgtga cccaaacatt gtctgggagc gagcgcgag ctttcgagcg caccgtgggtg 240
gcatcagcg cagtcacaaat ggggctgcaa atcaaccaacg tggttgacgg taaagctagc 300
tttcaactgc tccgcgctaa gtacgagzaa aaaaccgcca acaagaacaa atccgaacct 360
agcaggaggt acccaatcat gatag 385

```

<210> 30

<211> 12

<212> DNA

<213> Human immunodeficiency virus

<400> 30

atgggtgcga ga 12

<210> 31

<211> 12

<212> DNA

<213> Human immunodeficiency virus

<400> 31

gatgaggatt ag 12

<210> 32

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> gagpol-SYNgp

<400> 32

atgggcgccc gc 12

<210> 33

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> gagpol-SYNgp

<400> 33

gatgaggatt ag 12

<210> 34

<211> 12

<212> DNA

<213> Human immunodeficiency virus

<400> 34
atgagagtga ag 12

<210> 35
<211> 12
<212> DNA
<213> Human immunodeficiency virus

<400> 35
gctttgctat aa 12

<210> 36
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> synGP160mn

<400> 36
atgaggggtga ag 12

<210> 37
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> synGP160mn

<400> 37
gcgctgctgt aa 12

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [EP0368892A \[0009\]](#)
- [WO9712622A \[0009\]](#)
- [WO0055341A \[0013\]](#)
- [WO9941397A \[0013\] \[0015\] \[0024\]](#)
- [WO0015819A \[0014\]](#)
- [WO9932646A \[0027\] \[0028\] \[0204\] \[0205\] \[0206\] \[0213\]](#)
- [WO9915684A \[0075\]](#)
- [WO9205266A \[0113\]](#)
- [WO9805759A \[0122\]](#)
- [WO9805754A \[0122\]](#)
- [WO9717457A \[0122\]](#)
- [WO9609400A \[0122\]](#)
- [WO9100047A \[0122\]](#)
- [WO9915683A \[0143\]](#)

- [WO9805635A \[0151\]](#)
- [WO9807859A \[0152\]](#)
- [WO9809985A \[0153\]](#)
- [WO9851610A \[0205\] \[0223\]](#)
- [GB0003837W \[0218\]](#)
- [GB0009760A \[0227\]](#)

Non-patent literature cited in the description

- ZUFFEREY R et al. *Journal of Virology*, 1998, 9873-9880 [0014]
- MILLER, N.J. WHELAN *Hum Gene Ther.*, 1997, vol. 8, 803-15 [0226]
- LEWIS ERMAN J. *Virol.*, 1993, vol. 68, 510- [0226]
- NALDINI, LU. BLOMER P. GALLAY D. ORY R. MULLIGAN F. H. GAGE 1. M. VERMAD. TRONO *Science*, 1996, vol. 272, 263-7 [0226]
- MORGENSTERN LAND *Nucleic Acids Res.*, 1990, vol. 18, 3587-3596 [0226]
- CHANG, D. D. P. A. SHARP *Cell*, 1989, vol. 59, 789-795 [0226]
- WANG SEMENZ *Proc Natl Acad Sci.*, 1993, vol. 90, 430- [0226]
- DACHS et al. *Nature Med.*, 1997, vol. 5, 515- [0226]
- FIRTH et al. *Proc Natl Acad Sci.*, 1994, vol. 90, 6496-6500 [0226]
- MADAN et al. *Proc Natl Acad Sci.*, 1993, vol. 90, 3928- [0226]
- SEMENZ WANG *Mol Cell Biol.*, 1992, vol. 12, 5447-5454 [0226]
- TAKENAKA et al. *J Biol Chem.*, 1989, vol. 264, 2363-2367 [0226]
- PESHAVARIADAY *Biochem J.*, 1991, vol. 275, 427-433 [0226]
- INOU et al. *J Biol Chem.*, 1989, vol. 264, 14954-14959 [0226]
- OVERELL et al. *Mol Cell Biol.*, 1988, vol. 8, 1803-1808 [0226]
- ATTENELLO E *Science*, 1984, vol. 226, 187-190 [0226]
- GAZIT et al. *Cancer Res.*, 1985, vol. 55, 1660-1663 [0226]
- YU et al. *Proc Natl Acad Sci.*, 1986, vol. 83, 3194-3198 [0226]
- DOUGHERTY TEMIN *Proc Natl Acad Sci.*, 1987, vol. 84, 1197-1201 [0226]
- HAWLEY et al. *Proc Natl Acad Sci.*, 1987, vol. 84, 2406-2410 [0226]
- YEE, J. K. A. MIYANO HARAP. LAPORTEK. BOUIC J. C. BURNST. FRIEDMANN *Proc. Natl. Acad. Sci. USA.*, 1994, vol. 91, 9564-8 [0226]
- JOLLEY et al. *Nucleic Acids Res.*, 1983, vol. 11, 1855-1872 [0226]
- EMER MANTENIM *Cell*, 1984, vol. 39, 449-467 [0226]
- HERMAN COFFIN *Science*, 1987, vol. 236, 845-848 [0226]
- BENDER et al. *J Virol*, 1987, vol. 61, 1639-1646 [0226]
- PEAR et al. *Proc Natl Acad Sci*, 1993, vol. 90, 8392-8396 [0226]
- DANOS MULLIGAN *Proc Natl Acad Sci.*, 1998, vol. 85, 6460-6464 [0226]
- MARKOWITZ et al. *Virology*, 1988, vol. 167, 400-406 [0226]
- COSSET et al. *J. Virol.*, 1995, vol. 69, 7430-7436 [0226]
- MEBATSION et al. *Cell*, 1997, vol. 90, 841-847 [0226]
- MARSHALL *Nature Biotechnology*, 1998, vol. 16, 129- [0226]
- *Nature Biotechnology*, 1996, vol. 14, 556- [0226]
- WOLFF TRUBETSKOY *Nature Biotechnology*, 1998, vol. 16, 421- [0226]
- DUBRIDGE, R. B. P. TANGH. C. HSIAP.-M. LEONG J. H. MILLER M. P. CALOS *Mol. Cell. Biol.*, 1987, vol. 7, 379-387 [0226]
- GEY, G. O. W. D. COFFMAN M. T. KUBICEK *Cancer res.*, 1952, vol. 12, 264- [0226]
- KIM, S. Y. R. BYRN J. GROOPMAN D. BALTIMORE J. *Virol.*, 1989, vol. 63, 3708-3713 [0226]
- ADACHI, A. H. GENDELMANS. KOENIGT. FOLKS R. WILLEA. RABSON M. MARTIN J. *Virol.*, 1986, vol. 59, 284-291 [0226]
- HAAS, J. E.-C. PARK B. SEED *Current Biology*, 1996, vol. 6, 315- [0226]
- ZOLOTUKHIN, S. M. POTTER W. W. HAUSWIRTH J. GUYN. MUZYCZKAA "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells *J Virol.*, 1996, vol. 70, 4646-54 [0226]
- FISHER, A. E. COLLAL TIL. RATNERR. GALLOF. WONG-STAAL *Nature*, 1985, vol. 316, 262-265 [0226]
- KOZAK, M. *Annu. Rev. Cell Biol.*, 1992, vol. 8, 197-225 [0226]

- CASSAN, M.N. DELAUNAYC. VAQUEROJ. P. ROUSSETJ. Virol., 1994, vol. 68, 1501-8 [0226]
- PARKIN, N. T.M. CHAMORROH. E. VARMUSJ. Virol., 1992, vol. 66, 5147-51 [0226]
- J. VIROL., vol. 68, 3888-3895 [0226]
- MITROPHANOUS KYOON SROHLL JPATIL DWILKES FKIM VKINGSMAN SKINGSMAN AMAZARAKIS NGene Ther., 1999, vol. 6, 111808-18 [0226]
- ORY, D. S.B. A. NEUGEBORENR. C. MULLIGANProc Natl Acad Sci USA., 1996, vol. 93, 11400-6 [0226]
- ZHU, Z. H.S. S. CHENA. S. HUANGJ Acquir Immune Defic Syndr., 1990, vol. 3, 215-9 [0226]
- CHESEBRO, B.K. WEHRLYW. MAURYJ Virol., 1990, vol. 64, 4553-7 [0226]
- SPECTOR, D. H.E. WADED. A. WRIGHTV. KOVALC. CLARKD. JAQUISHS. A. SPECTORJ Virol., 1990, vol. 64, 2298-2308 [0226]
- VALSESIA WITTMANN, S.A. DRYNDAG. DELEAGEM. AUMAILLEYJ. M. HEARDO. DANOSG. VERDIERF. L. COSSET. J. Virol., 1994, vol. 68, 4609-19 [0226]
- KIM, V. N.K. MITROPHANOUS. M. KINGSMANK. A. J.J Virol, 1998, vol. 72, 811-816 [0226]
- SONEOKA, Y.P. M. CANNONE. E. RAMSDALEJ. C. GRIFFITHSG. ROMANOS. M. KINGSMANA. J. KINGSMANNucleic Acids Res., 1995, vol. 23, 628-33 [0226]
- RNA blot analysisSAGERSTROM, C.H. SIVEA laboratory guide to RNA: isolation, analysis and synthesisWiley-Liss Inc.19960000vol. 1, 83-104 [0226]
- MALIM, M. H.J. HAUBERS. Y. LEJ. V. MAIZELB. R. CULLENNature, 1989, vol. 338, 254-7 [0226]
- FELBER, B. K.M. HADZOPOULOU CLADARASC. CLADARAST. COPELAND G. N. PAVLAKISProc. Natl. Acad. Sci. USA., 1989, vol. 86, 1495-1499 [0226]
- HADZOPOULOU CLADARAS, MB. K. FELBERC. CLADARASA. ATHANASSOPOULOSA. TSEG. N. PAVLAKISJ. Virol., 1989, vol. 63, 1265-74 [0226]
- SCHNEIDER, R.M. CAMPBELLG. NASIOULASB. K. FELBERG. N. PAVLAKISJ Virol., 1997, vol. 71, 4892-903 [0226]
- SCHWARTZ, S.B. K. FELBERG. N. PAVLAKISJ. Virol., 1992, vol. 66, 150-159 [0226]
- MALDARELLI, F.M. A. MARTINK. STREBELJ. Virol., 1991, vol. 65, 5732-5743 [0226]
- MIKAELIAN, I.M. KRIEGM. GAITJ. KARNJ. Mol. Biol., 1996, vol. 257, 246-264 [0226]
- XU, N.C.-Y. CHENA.-B. SHYUMol. Cell. Biol., 1997, vol. 17, 4611-4621 [0226]
- NALDINI, L.Curr. Opin. Biotechnol., 1998, vol. 9, 457-463 [0226]
- PAROLIN, C.T. DORFMANG. PALUH. GOTTLINGERJ. SODROSKIJ. Virol., 1994, [0226]
- DULL, T.R. ZUFFEREYM. KELLYR. MANDELM. NGUYEND. TRONOL. NALDINIJ. Virol., 1998, vol. 72, 8463-8471 [0226]
- CUI, Y.T. IWAKAMAL.-J. CHANGJ. Virol., 1999, vol. 73, 6171-6176 [0226]
- CHANG, L.-J.V. URLACHERT. IWAKAMAY. CUIJ. ZUCALI.Gene Ther, 1999, vol. 6, 715-728 [0226]
- HARRISON, G.G. MIELEE. HUNTERA. LEVERJ. Virol., 1998, vol. 72, 5886-5896 [0226]
- MCBRIDE, M. S.A. T. PANGANIBANJ. Virol., 1997, vol. 71, 2050-8 [0226]
- LEVER, A.H. GOTTLINGERW. HASELTINEJ. SODROSKIJ. Virol., 1989, vol. 63, 4085-7 [0226]
- BRIGHTY, D.M. ROSENBERGProc. Natl. Acad. Sci. USA., 1994, vol. 91, 8314-8318 [0226]
- ARAI, T.M. TAKADAM. UIH. IBAVirology, 1999, vol. 260, 109-115 [0226]
- FORNEROD, M.M. OHNOM. YOSHIDAI. W. MATTAJCell, 1997, vol. 90, 1051-1060 [0226]
- FRIDELL, R. A.H. P. BOGERDB. R. CULLENProc. Natl. Acad. Sci. U S A., 1996, vol. 93, 4421-4 [0226]
- POLLARD, V.M. MALIMAnnu. Rev. Microbiol., 1998, vol. 52, 491-532 [0226]
- STADE, K.C. S. FORDC. GUTHRIEK. WEISCell, 1997, vol. 90, 1041-1050 [0226]
- ULLMAN, K. S.M. POWERS, AD. J. FORBESCell, 1997, vol. 90, 967-970 [0226]
- OTERO, G. C.M. E. HARRISJ. E. DONELLOT. J. HOPEJ. Virol., 1998, vol. 72, 7593-7597 [0226]
- WOLFF et al.Chem Biol., 1997, vol. 4, 139-147 [0226]

Patentkrav

1. Fremgangsmåde til frembringelse af en replikationsdefekt HIV-vektor, som omfatter transficering af en produktionscelle med følgende:

i) et HIV-genom, der omfatter et nukleotid af interesse (NOI), et pakningssignal og RRE;

ii) en nukleotidsekvens, der koder for HIV-gag- og pol-proteiner; og

iii) en nukleotidsekvens, der koder for et env-protein; der er kendetegnet ved, at nukleotidsekvensen, der koder for HIV-gag- og pol-proteinerne, er kodon-optimeret til ekspression i produktionscellen med undtagelse af sekvensen, der omslutter læserammeforskydningsstedet, og hvor (ii) ikke omfatter RRE; hvor produktionscellen udtrykker rev, og hvor HIV-vektoren er en selv-inaktiverende vektor.

2. Fremgangsmåde ifølge krav 1, hvor den kodon-optimerede nukleotidsekvens har sekvensen, der er vist i SEQ ID NO: 15.

3. HIV-vektorpartikel-produktionscelle, der omfatter:

i) et HIV-genom, der omfatter mindst én nukleotidsekvens af interesse, et pakningssignal og RRE; og

ii) en nukleotidsekvens, der koder for HIV-gag- og pol-proteiner; og

iii) en nukleotidsekvens, der koder for et env-protein; der er kendetegnet ved, at nukleotidsekvensen, der koder for HIV-gag- og pol-proteinerne, er kodon-optimeret til ekspression i produktionscellen med undtagelse af sekvensen, der omslutter læserammeforskydningsstedet, og hvor (ii) ikke omfatter RRE; hvor produktionscellen udtrykker rev, og hvor HIV-vektoren er en selv-inaktiverende vektor.

4. Produktionscelle ifølge krav 3, hvor den kodon-optimerede nukleotidsekvens har sekvensen, der er vist i SEQ ID NO: 15.

5. Produktionscelle ifølge krav 3, hvor kappegenet er kodon-optimeret.

6. Produktionscelle ifølge et hvilket som helst af kravene 3 til 5, hvor nukleotidet af interesse er udvalgt blandt et terapeutisk gen, et markørgen og et selektionsgen.

5

7. Produktionscelle ifølge et hvilket som helst af kravene 3 til 6 til anvendelse i en fremgangsmåde til frembringelse af HIV-vektorpartikler.

10 8. Farmaceutisk sammensætning, der omfatter produktionscellen ifølge et hvilket som helst af kravene 3 til 7 sammen med et farmaceutisk acceptabelt bæremateriale eller fortyndingsmiddel.

DRAWINGS

FIGURE 1

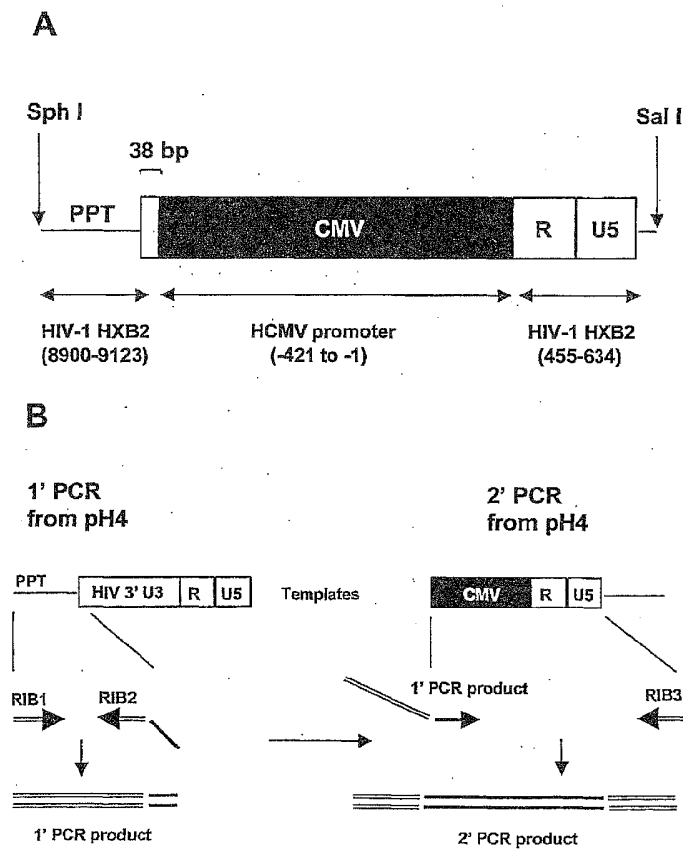


FIGURE 2

gappol-HXB2 -> Codon Usage

DNA sequence 4308 b.p. ATGGGTGCGAGA ... GATGAGGATTAG linear

1436 codons

MW : 161929 Dalton CAI(S.c.) : 0.083 CAI(E.c.) : 0.151

TTT phe F	21	TCT ser S	3	TAT tyr Y	30	TGT cys C	18
TTC phe F	14	TCC ser S	3	TAC tyr Y	9	TGC cys C	2
TTA leu L	46	TCA ser S	19	TAA och Z	-	TGA opa Z	-
TTG leu L	11	TCG ser S	1	TAG amb Z	1	TGG trp W	37
CTT leu L	13	CCT pro P	21	CAT his H	20	CGT arg R	-
CTC leu L	7	CCC pro P	14	CAC his H	7	CGC arg R	-
CTA leu L	17	CCA pro P	41	CAA gln Q	56	CGA arg R	3
CTG leu L	16	CCG pro P	-	CAG gln Q	39	CGG arg R	3
ATT ile I	30	ACT thr T	24	AAT asn N	42	AGT ser S	18
ATC ile I	14	ACC thr T	20	AAC asn N	16	AGC ser S	16
ATA ile I	56	ACA thr T	43	AAA lys K	88	AGA arg R	45
ATG met M	29	ACG thr T	1	AAG lys K	34	AGG arg R	18
GTG val V	15	GCT ala A	17	GAT asp D	37	GGT gly G	11
GTC val V	11	GCC ala A	19	GAC asp D	26	GGC gly G	10
GTA val V	55	GCA ala A	55	GAA glu E	75	GGA gly G	61
GTG val V	15	GCG ala A	5	GAG glu E	32	GGG gly G	26

FIGURE 3a

gagpol-SYNgp [1 to 4308] -> Codon Usage

DNA sequence 4308 b.p. ATGGGCCCCCGC ... GATGAGGATTAG linear

1436 codons

MW : 161929 Dalton CAI(S.c.) : 0.030 CAI(E.c.) : 0.296

TCT phe F	5	TCT ser S	5	TAT tyr Y	10	TGT cys C	6
TTC phe F	30	TCC ser S	11	TAC tyr Y	29	TGC cys C	14
TTA leu L	2	TCA ser S	4	TAA och Z	-	TGA opa Z	-
TTG leu L	7	TCG ser S	6	TAG amb Z	.1	TGG trp W	37
CTT leu L	3	CCT pro P	14	CAT his H	6	CGT arg R	2
CTC leu L	22	CCC pro P	39	CAC his H	21	CGC arg R	34
CTA leu L	6	CCA pro P	10	CAA gln Q	14	CGA arg R	3
CTG leu L	70	CCG pro P	13	CAG gln Q	81	CGG arg R	10
ATT ile I	17	ACT thr T	11	AAT asn N	13	AGT ser S	7
ATC ile I	79	ACC thr T	48	AAC asn N	45	AGC ser S	27
AUA ile I	4	ACA thr T	13	AAA lys K	25	AGA arg R	7
ATG met M	29	ACG thr T	16	AAG lys K	97	AGG arg R	13
GTT val V	5	GCT ala A	15	GAT asp D	19	GTY gly G	10
GTC val V	27	GCC ala A	56	GAC asp D	44	GGC gly G	54
GTA val V	6	GCA ala A	13	GAA glu E	29	GGA gly G	16
GTG val V	58	GCG ala A	12	GAG glu E	78	GGG gly G	28

Codon usage in human genes (MH), wild type HIV-1 Gag-pol (WT) and the codon optimised HIV-1 Gag-pol (CO)

		MH	WT	CO			MH	WT	CO			MH	WT	CO			MH	WT	CO				
Ala	A	13	46	8	Cys	C	68	10	70	Leu	A	3	14	3	Ser	C	34	35	55				
	GC	C	53	19		65	TG	T	32		90	30	CT	C		26	8	17	AG	T	10	10	3
	G	17	11	8								G	58	14		70	TC	A	5	38	17		
	T	17	24	19								T	5	11		8		C	28	10	14		
Arg	A	10	58	10	Gln	A	12	53	21	TT	A	2	42	5		G	9	3	7				
	AG	G	18	29		11	CA	G	88		47	79	G	6		11	0	T	13	3	3		
	CG	A	6	6		0																	
	C	37	0	61																			
	G	21	6	10	Glu	A	25	65	38	Lys	A	18	58	28	Thr	A	14	45	16				
	T	7	0	5		GA	G	75	35		62	AA	G	32		42	72	AC	C	57	29	52	
																		G	15	0	19		
																		T	14	26	13		
Asn	C	78	29	71	Gly	A	14	53	21	Phe	C	80	45	45	Tyr	C	74	20	80				
	AA	T	22	71		29	GG	C	50		21	55	TT	T		20	55	55	TA	T	28	80	20
Asp	C	75	84	70	His	C	79	80	90	Pro	A	16	52	24	Val	A	5	56	4				
	GA	T	25	36		30	CA	T	21		70	10	CC	C		43	16	39	GT	C	25	8	20
					Ile	A	5	58	0							G	64	24	76				
						AT	C	18	19		92							T	7	12	0		
						T	77	23	0														

Figure 3b

FIGURE 4

env-mn [1 to 2571] -> Codon Usage

DNA sequence 2571 b.p. ATGAGAGTGAAG ... GCTTTGCTATAA linear

857 codons

MW : 97078 Dalton CAI(S.c.) : 0.083 CAI(E.c.) : 0.140

TTT phe F	13	TCT ser S	7	TAT tyr Y	15	TGT cys C	16
TTC phe F	11	TCC ser S	3	TAC tyr Y	7	TGC cys C	5
TTA leu L	20	TCA ser S	13	TAA och Z	1	TGA opa Z	-
TTG leu L	17	TCG ser S	2	TAG amb Z	-	TGG trp W	30
CTT leu L	9	CCT pro P	5	CAT his H	8	CGT arg R	-
CTC leu L	11	CCC pro P	9	CAC his H	6	CGC arg R	2
CTA leu L	12	CCA pro P	12	CMA gln Q	22	CGA arg R	1
CTG leu L	15	CCG pro P	2	CAG gln Q	19	CGG arg R	1
ATT ile I	21	ACT thr T	16	AAT asn N	50	AGT ser S	18
ATC ile I	10	ACC thr T	14	AAC asn N	23	AGC ser S	11
ATA ile I	32	ACA thr T	28	AAA lys K	32	AGA arg R	30
ATG met M	17	ACG thr T	5	AAG lys K	14	AGG arg R	15
GTT val V	8	GCT ala A	16	GAT asp D	18	GGT gly G	10
GTC val V	9	GCC ala A	7	GAC asp D	14	GGC gly G	6
GTA val V	26	GCA ala A	20	GAA glu E	36	GGA gly G	28
GTG val V	12	GCG ala A	5	GAG glu E	10	GGG gly G	12

FIGURE 5

SYNBP150mm -> Codon Usage

DNA sequence 2571 b.p. ATGAGGTTGAAG ... GCGCTGCTGTAA linear

857 codons

MW : 97078 Dalton CAI(S.c.) : 0.074 CAI(E.c.) : 0.419

TTC phe F	-	TCT ser S	2	TAT tyr Y	1	TGT cys C	-
TTC phe F	24	TCC ser S	4	TAC tyr Y	21	TGC cys C	21
TTA leu L	-	TCA ser S	-	TAA och Z	1	TGA opa Z	-
TTG leu L	-	TCG ser S	-	TAG amb Z	-	TGG trp W	30
CTT leu L	-	CCT pro P	-	CAT his H	2	CGT arg R	1
CTC leu L	20	CCC pro P	26	CAC his H	12	CGC arg R	36
CTA leu L	1	CCA pro P	-	CAA gln Q	-	CGA arg R	-
CTG leu L	63	CCG pro P	2	CAG gln Q	41	CGG arg R	4
ATT ile I	2	ACT thr T	-	AAT asn N	2	AGT ser S	-
ATC ile I	61	ACC thr T	59	AAC asn N	61	AGC ser S	48
ATA ile I	-	ACA thr T	-	AAA lys K	1	AGA arg R	2
ATG met M	17	ACG thr T	4	AAG lys K	45	AGG arg R	6
GTT val V	-	GCT ala A	-	GAT asp D	2	GGT gly G	1
GTC val V	1	GCC ala A	40	GAC asp D	30	GGC gly G	47
GTA val V	1	GCA ala A	-	GAA glu E	3	GGA gly G	-
GTG val V	53	GCG ala A	8	GAG glu E	43	GGG gly G	8

FIGURE 6
HIV Constructs

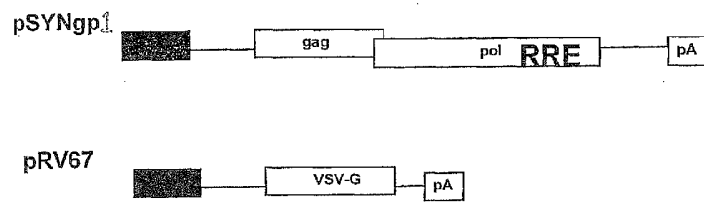
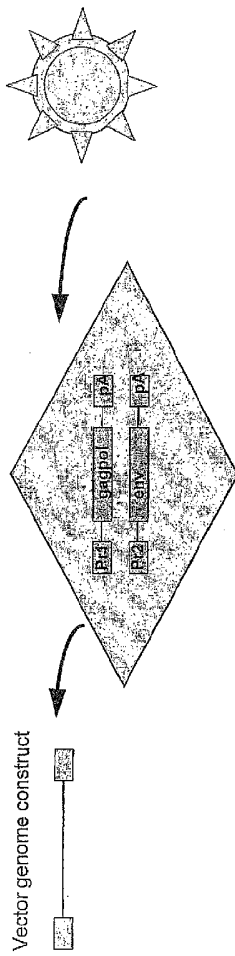


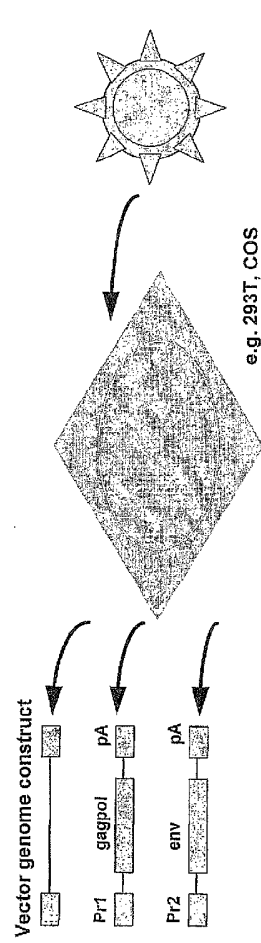
FIGURE 7

The Hit Vector System

Helper packaging cell lines



Three-plasmid cotransfection (HIT)



[illegible]

[illegible]

4201 ATATGCACTCATAGTGTGGCCACGAAGCAGAAAGCGAAGCATTTAGGCAT 4250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4201 ATATGCACTCATAGTGTGGCCACGAAGCAGAAAGCGAAGCATTTAGGCAT 4250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4251 TATGCGCAACAGATGTCGCGTGATTCCTTGGCGAGCACAGATAGA 4300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4251 TTGTGAAACAAGATGTCGCGTGATTCCTTGGCGAGCACAGATAGA 4300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4301 GCGATTAG 4307
||||| |||||
4301 GCGATTAG 4307
||||| |||||

3601 TGGACTTAGCCCCCGCATCTGGCACTCCCATGCCACCACATCTGAGGC 3650
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3601 TAGACTTGATTCACGGAATTTGGCACTTAAGATTTGATCAATTTAGAGGA 3650
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3651 AAGGTTATCTCGTGTCGCGCTCCATCTGGCCACATGTCATCATCGAGCGCA 3700
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3651 AAAGTTATCTCGTGTCGCGCTCCATCTGGCCACATGTCATCATCGAGCGCA 3700
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3701 GTCTATCTCCGCGCAACAGCGGCGAGCGAGCAATCTACTCTCTCGAGCGC 3750
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3701 AGTATATCTCCGCGCAACAGCGGCGAGCGAGCAATCTACTCTCTCGAGCGC 3750
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3751 TSCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3800
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3751 TACACAGAGATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3800
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3801 TTCACACAGTGTCTAGGTTTAGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3801 TTTCACACAGTGTCTAGGTTTAGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3850
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3851 GGAGTTTGAGTATCCCTTACACTCCCAGAGTCAAGGCGCTGCTGCGAGCTCA 3900
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3851 GGAATTTGAGTATCCCTTACACTCCCAGAGTCAAGGCGCTGCTGCGAGCTCA 3900
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3901 TGAATAAGCGAGTTTAAAGAAATATTCCGCGCTGCTGCGAGAGTCAAGGCTGAG 3950
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3901 TGAATAAGCGAGTTTAAAGAAATATTCCGCGCTGCTGCGAGAGTCAAGGCTGAG 3950
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3951 CATCTCAAGACCGCTGCTCCAATGGCGCTATTATCATCAATTTGAGCG 4000
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3951 CATCTTAAAGCGAGCTGACAAATGGCGCTATTATCATCAATTTGAGCG 4000
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4001 CAGCGGCGGCTATGCGCGCTGACATCTCGCGGCGAGCGAGATCTGAGCATCA 4050
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4001 AAAAGGCGGCTATGCGCGCTGACATCTCGCGGCGAGCGAGATCTGAGCATCA 4050
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4051 TCGCGAGCGCTCATCGAGCTTAGGAGCTGCAAAAACAGATTTACCGAGATT 4100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4051 TAGCGAGCGCTCATCGAGCTTAGGAGCTGCAAAAACAGATTTACCGAGATT 4100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4101 CGAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4101 CAATATTTCCGCTTTATACAGGCGAGCGAGATATCACTTTGAGGAG 4150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4151 CCACGAGAGCTCTCTGCTGAGNGGTGAGGGCGCTAGTGTGTCAGAGCTA 4200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4151 ACAGAGAGCTCTCTGCTGAGNGGTGAGGGCGCTAGTGTGTCAGAGCTA 4200
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

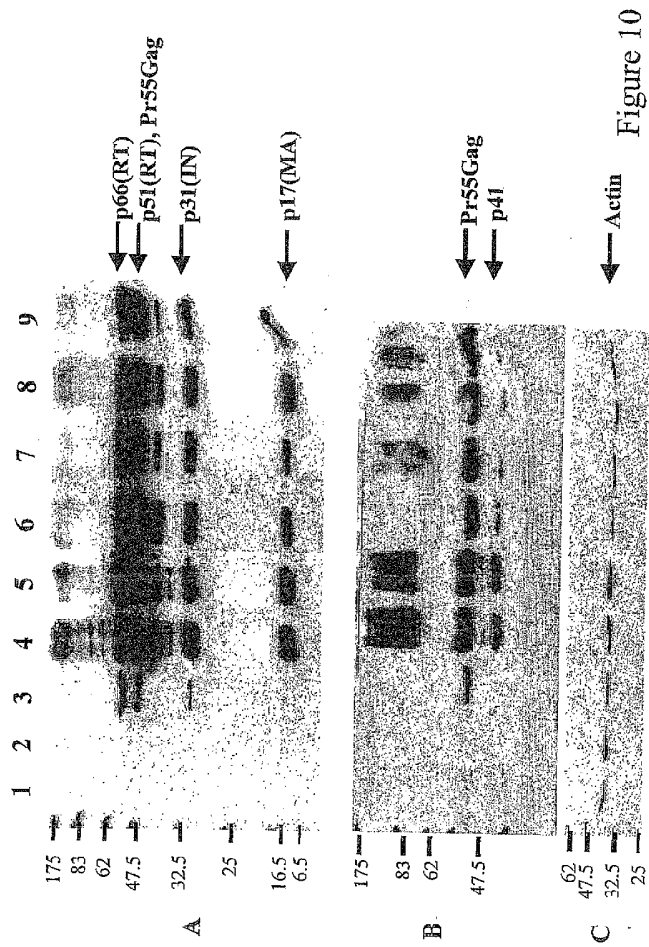
WT	ATGGGAGACCCCTTGACATGGAGCAAGGCGCTCAAGAAGTTAGAGAAGGTGACGGTACAA	60
CO	ATGGGCGATGCCCTCACCTGGTCCAAGCCCTGARGAAACTGGAAAAGTCACCGTTCAG	60
WT	GGGTCTCAGAAAATTAACTACTGGTAACTGTAAATTGGGCGCTAAGTCTAGTAGACTTTATTT	120
CO	GGTAGCCAAAAGCTTACCACAGGCRAATGCAACTGGGCAITGTCCCTGGTGGATCTTTTC	120
WT	CATGATACCAACTTTGTAAAAGAAAAGGACTGGCAGCTGAGGGATGTCATTCCATTGCTG	180
CO	CACGACACTAATTTCTTTAAGGAGAAAGATTGGCACTCAGAGACGTGATCCCCCTCTTG	180
WT	GAAGATGTAACCTCAGACGCTGTCAAGACAGAAAGAGAGGCCCTTTGAAAAGACATGTGCG	240
CO	GAGGACGTGACCCAAACATTGTCTGGGCAAGAGCGCGAAGCTTTGAGCGCACCTGTGTG	240
WT	GCAATTTCTGCTGTAAAGATGGGCGCTCCAGATTAAATAATGTAGTAGATGGAAAGGCATCA	300
CO	GCCATCAGCGAGTCAAATGGGGCTGCAATCAACCACTGTGTTGACGCTAAGCTAGC	300
WT	TTCCAGCTCCTAAGAGCGAAATATGAAAAGAGACTGCTAATAAAAAGCAGTCTGAGGCC	360
CO	TTTCAACTGCTCCGCGCTAAGTACGAGAAAGAAACCGCCACAGAAACATCCGAACCT	360
WT	TCTGAGGAATATCCAATCATGATAGATGGGCTGGAACAGAAATTTAGACCTTAACA	420
CO	AGCGAGGAGTACCCCAATTATGATCGAGCGCGCGCAATAGGAACTTCGCGCACTGACT	420
WT	CCTAGAGGATATCTACTTTGGGTGAATACCATACAGACAAATGGTCTATTAAATGAGCT	480
CO	CCAGAGGGCTATACCACTGGGTCAACACCATCCAGACAAGCGACTTTTGAACGAAAGCC	480
WT	AGTCAAACTTATTTGGGATATTATCAGTAGACTGTACTTCTGAAGAAATGAATGCATT	540
CO	TCCAGAAACCTGTTCCGCACTCCGTCTCTGTGACTGCACTCCGAAGAAATGAATGCTTT	540
WT	TTGGATGTGTACCTGGCCAGGCAGGACAAAAGCAGATATTACTTGATGCAATTGATAAG	600
CO	CTCGACGTGGTGCAGGACAGGCTGGAAGAAAACAGATCCCTGCTCGATGCCATTGACAA	600
WT	ATAGCAGATGATTGGGATAATAGACATCCATTACGGAATGCTCCACTGGTGGCACCACCA	660
CO	ATCGCGACGACTGGGATAATCGCCACCCCTGCAAAACGCCCTCTGTGGCTCCCA	660
WT	CAAGGGCCCTATTCCCATGACAGCAGGTTTATTAGAGGTTTAGGAGTACCTAGAGAAAGA	720
CO	CAGGGGCCCTATCCCTATGACCGCTAGGTTTATTAGGGGACTGGGGGTGCCCGCGAAGC	720
WT	CAGATGGAGCCTGCTTTTGTATCAGTTTAGGCAGACATATAGACAATGGATAATAGAAGCC	780
CO	CAGATGGAGCAGCATTTGACCAATTTAGGCAGACCTACAGACAGTGGATCATCGAAGCC	780
WT	ATGTCAGAAGGCATCAAAGTGAATGGAAGAACCTAAAGCTCAAAATATTAGGCAAGGA	840
CO	ATGACCGAGGGGATTAAAGTCAATGATCGGAAGCCCAAGGCACAGAACATCAGGCAGGGG	840
WT	GCTAAGGAACCTTACCCAGAAATTGTAGACAGACTATTATCCCAATAAAAAGTGAGGGA	900
CO	GCCAGGAACCATACCTTGAATTTGTGACAGGCTTCTGTCCAGATTAAATCCGAAGGC	900
WT	CATCCACAAGAGATTTCAAAATTCCTTGAATGATACACTGACTATTTCAGAAACCAATGAG	960
CO	CACCCCTCAGGAGTCTCCAAGTCTTGAACAGACACACTGACTATTCAAAATGCAATGAA	960

WT	GAATCTAGAAATGCTATGAGACATTTAAGACUAGAGGATACATTAGAGAGAAATGTAT	1020
CO	GAGTGCAGAAACGCCATGAGGCACCTCAGACCTGAGAGATACCTTGGAGGAGAAATGTAT	1020
WT	GCCTTGCAAGACATTGGAACACAAAACAAAGATGATGTTATTTGGCAAAAGCACTTCAG	1080
CO	GCATGTGCGACATTTGGCACTACCAAGCAAAAGATGATGCTGCTCCCAAGGCTCTGCAA	1080
WT	ACTGGTCTTGGCGGCCCATTTAAAGGTGGAGCCTTGAAGGAGGGCCACTAAAGGCGCA	1140
CO	ACCGGCTTGGCTGGTCCATTCAAGGAGGAGCCTTGAAGGAGGGTCCATTGAAGGCTGCA	1140
WT	CAACATGTTATAACTGTGGGAAGCCAGGACCTTTATCTAGTCAATGTAGAGCACTTAAA	1200
CO	CAACATGTTATAAATTGTGGGAAGCCAGGACCTTTATCTAGTCAATGTAGAGCACTTAAA	1200
WT	GTCTGTTTAAATGTAAACAGCCTGGACATTTCTCAAGCAATSCAGAGTGTTCBAAA	1260
CO	GTCTGTTTAAATGTAAACAGCCTGGACATTTCTCAAGCAATSCAGAGTGTTCBAAA	1260
WT	AACGGGAAGCAAGGGCTCAAGGAGGCCCCAGAAACAACTTTCCCGATACACAGAAAG	1320
CO	AACGGGAAGCAAGGGCTCAAGGAGGCCCCAGAAACAACTTTCCCGATACACAGAAAG	1320
WT	AGTCAGCACACAAATCTGTTGTACAAGAGACTCCTCAGACTCAAAATCTGTACCCAGAT	1380
CO	AGTCAGCACACAAATCTGTTGTACAAGAGACTCCTCAGACTCAAAATCTGTACCCAGAT	1380
WT	CTGAGCGAAATAAAAAGGAATACAAAGTCAAGGTGAAGGATCAAGTACAGGATCTCAAC	1440
CO	CTGAGCGAAATAAAAAGGAATACAAAGTCAAGGTGAAGGATCAAGTACAGGATCTCAAC	1440
WT	CTGGACAGTTTGTGGGAGTAACTATATCTAGAGAAAGGCCCTACTCAATAGTATTAA	1500
CO	CTGGACAGTTTGTGGGAGTAACTATATCTAGAGAAAGGCCCTACTCAATAGTATTAA	1500
WT	TTAATGATACTCCCTTAAATGTACTGTAGACACAGGAGCAGATACTTCAGTCTTGACTA	1560
CO	TCAATGACRCCCTCTTAATGTCTGTGGACRCCGAGCCGACACCGGCTTCTCACTA	1560
WT	CTGCACATTATAATAGGTTAAATATAGAGGGAGAAATATCAAGGGACGGGAATATAG	1620
CO	CTGCTCACTATAACGAGCTGAATATACAGAGGAAGGAATATCAAGGGACGGGAATATAG	1620
WT	GAGTGGGAGGAAATGTGAAACATTTTCTACGCTGTGACTATAAGGAAAGGGTAGAC	1680
CO	GCGTTGAGGCAACGTGGAACCTTTTCCACTCTGTCACTCAAAAGAGGGGAGAC	1680
WT	ACATTAGACAAAGATGCTAGTGGCAGATATTCAGTGACTATTTTGGGACGAGATATTC	1740
CO	ACATTAAACCAAGATGCTGGTGGCGGACATCCCGGTCACTTCTTGGCAGAGCAATTC	1740
WT	TTGAGCACTTAGGTGCAAAATGGTTTGGCAGGCTCTCAAGGAAATAAAAATTAGAA	1800
CO	TCCAGGACTTGGGCGCTRAACTCGTGTGGCAGACTGTCTAAGGAATCAAGTTCCGCA	1800
WT	AAATAGAGTTAAAGAGGGCCACATGGGGCCAAATATCTCTAATGGCCACTCACTAAGG	1860
CO	AGATCGAGCTGAAGAGGGCCACATGGGTCCAAATATCCCTAGTGGCCCTGACCTAAGG	1860
WT	AGAACTAGAAAGGGGCCAAGAGATAGTCCAAAGACTATTTCTAGAGGGCAAAATATCAG	1920
CO	AGAACTAGAGGGCGCTAAGGAATCGTGCAGGCGCTGCTTCTAGAGGGCAAGATTAGCG	1920
WT	AAGCTAGTGACATTAATCCTTATAATTCACCCATATTTGTAATAAAAAGAGGTCTGCA	1980
CO	AGGCCAGGACATTAATCCTTACAAAGCCCATCTTTGTGATTAGAAAAAGAGGCGCA	1980

WT	AATGGAGGTTATTACAAGATCTGAGAGAATTAAACRAAACAGTACAGTAGHAACGGAAA	2040
CO	AATGGAGAGCTCCGCGAGGACCTGAGGGAACTCAACRAGACCGTCCAGGTGGAACCTGAGA	2040
WT	TATCCAGAGGATTGCGCTCACCGGGAGGATTAAATTAATGTAAACACATGACTGTATTAG	2100
CO	TCTCTCGCGGACGCTTACCUCGCGCGCTGATTAAATGCAAGCACATGACAGTCCCTTG	2100
WT	ATAITGGAGATGCATATTTCACTATACCTTTAGATCCAGAGTTAGACCATATACAGCTT	2160
CO	ACATTGGAGAGCGCTTATTTTACCATCCCCCTCGATCCTGAATTTGCGCCCTATACGTCTT	2160
WT	TCACATATCCCTCCATTAAATCATCAAGAACAGATAAAAGATATGTGTGGAAATGTTTAC	2220
CO	TTACCATCCCCAGCATCAATCACCGGAGCCCGATAAACGCTATGTGTGGAGTGCCTCC	2220
WT	CACAAAGGATTGCTGTGAGCCCCATATATATATCAGAAAACATTACAGGAAATTTTACAAC	2280
CO	CCCAGGGAATTTGTCTTAGCCCCCTACATTTTACAGAAAGACCTTCAAGAGATCCTCCAAC	2280
WT	CTTTTAGGGAAGATATCCTGAAGTACAATTGTATCAATATATGGAGTATTTGTTTCATGG	2340
CO	CTTTCCCGGAAGATACCCAGAGGTTCAACTCTACCAATATATGGACGACCTGTTTCATGG	2340
WT	GAGTAATGGTTCTAAAAACAACAACAAGAGTTAATCATAGAAATTAAAGGCGATCTTAC	2400
CO	GGTCCACCGGGCTTAAGAAGCAGCAAGGAACTCATCTTCGAAGTGGGGCACTCTCC	2400
WT	TGAAAAAGGGTTTTGAGACACCCAGATGATAATTACAGAAGTGCCACCTTATAGCTGGC	2460
CO	TGGAGAAAGGCTTCGAGACACCCGACGACAGCTGCAAGAAGTTCCCTCCATATAGCTGGC	2460
WT	TAGGTTATCAACTTTGCTGAAATTTGGAAGTACAAAATGCAATTAGACATGGTAA	2520
CO	TGGCTTACAGCTTTGCGCTGAAACTGGAAAGTCCAGAGATGCAAGTTGGATATGGTCA	2520
WT	AGAAATCCAACTTAATGATGCAAAAATTAATGGGGAATATAACATGGATGAGCTCAG	2580
CO	AGAACCAACACATGAACGAGCTCCAGAGCTCATGGGCATATATTACCTGGATGAGCTCCG	2580
WT	GGATCCCAAGGTTGACAGTAAACACATTCAGCTACTACTAAGGGATGTTTAGAGTTGA	2640
CO	GAATCCCTGGGCTTACCGTTAAGCACATTCGCGCACTACAAAAGGATGCCCTGGAGTTGA	2640
WT	ATCAAAAAGTAATTTGGACGGAAGAGGCACAAAAGAGTTAGAAGAAAATTAATGAGAAGA	2700
CO	ACCAGAAGGTCTATTTGGACAGAGGAGCTCAGAAGGAACTGGAGGAGATAATGAAAAGA	2700
WT	TTAAAAATGCTCAAGGGTTACAATATTATAATCCAGAAGAGAAATGTTATGTTGAGGTTG	2760
CO	TTAAGAATGCTCAAGGGCTCCAACTACTACAATCCGAGAGAGAAATGTTGTTGCGAGGTCG	2760
WT	AAATACAAAAAATTAAGAGGCAACTTATGTTATAAAACAATCACAAGGAATCTTATGGG	2820
CO	AAATCACTAAGAACTACGAAGCCACCTATGTATCAACAGTCCCAAGGCATCTTGTGGG	2820
WT	CAGGTAAAGAGATTATGAAGGCTAATAAGGGATGTTCAACAGTAAAAAATTTAATGTTAT	2880
CO	CCGGAAGAGAAATCATGAAGGCCAACAAAGGCTGTTCCACCGTTAAAAATCTGATGCTCC	2880
WT	TGTTGCAACTGTGGCAACAGAAAGTATTACTAGAGTAGGAAAATGTCACAGTTTAAAGC	2940
CO	TGCTCCAGCACGTGGCCACCGAGTCTATCACCGGCTGGCAAGTGCCCCACCTTCAAAG	2940
WT	TACCATTTTACCAAGAGCAACTAATGTGGGAAATGCAAAAAGGATGGTATTAATCTTGGC	3000
CO	TTCCCTTCACTAAGGAGCAGGTGATGTGGAGATGCAAAAAGGCTGTTACTACTCTTGGC	3000

WT	TCCGAGAAATAGTATATACACATCAAGTAGTTTATGATGATTGGAGAATGAAATTTGGTAG	3060
CO	TTCCCGAGATCGTCTACACCCACCAAGTGGTGCACGACGACTGGAGAATGAAGCTTGTCTG	3060
WT	AAGAACCTACATCAGGAATAACAATATACACTGATGGGGGAAAAAATGGAGAGGAA	3120
CO	AGGAGCCCACTAGCGGAATTACAATCTATACCGACGGCGGAAGCAAAACGGAGAGGGAA	3120
WT	TAGCAGCTTATGTGACCAAGTAATGGGAGAACTAAACGAAAGGTTAGGACCTGTCACTC	3180
CO	TGCGTCATAGCTCACAATCTAACGGCCGCACCAAGCAAAGAGGCTCGGCCCTGTCACTC	3180
WT	ATCAAGTTGCTGAAAGAATGGCAATACAATGGCATTAGAGGTACCGAGATAAACAG	3240
CO	ACCAGGTGGCTGAGAGGATGGCTATCCGATGGCCCTTGAGGACACTAGAGACAAGCAGG	3240
WT	TAAATATAGTAACGTATAGTATTTATTTGTTGGAAAAATATTACAGAAGGATTAGGTTTAC	3300
CO	TGAACATTTGTAGCTGACAGCTACTCTGCTGGAAAAACATCACGAGGGCCTTGGCCCTGG	3300
WT	AAGGACCACAAAGTCCTTGGTGGCCCTATATACAAAATAACGAGAAAAAGAGATAGTTT	3360
CO	AGGAGCCCACTCTCCCTGGTGGCCCTATCATCCAGATATCCGCGAAAGGAATTTGTCT	3360
WT	ATTTTGTCTGGGTACCTGGTCAAAAGGGATATATGGTAATCAATTGGCAGATGAAGCG	3420
CO	ATTTTCGCTTGGGTGCTTGGACAAAGGAATTTACGGCAACCAACTCGCCGATGAAGCG	3420
WT	CAAAATATAAAGAGAGATCATGCTAGCATACCAAGGCACACAAATTAAGAGGAAAGAG	3480
CO	CCAAATTTAAGAGGAAATCATGCTTGCCCTACAGGGCACACAGATTAAAGGAGAGAGAG	3480
WT	ATGAAGATCGAGGGTTTGACTTATGTGTTCCCTATGACATCATGATACCTGTATCTGACA	3540
CO	ACGAGGACCGCTGGCTTTGACCTGTGTGTGCCMTACGACATCAFGATTCCCTTAGCGACA	3540
WT	CAAAATCATACCCACAGATGTAAAAATTCAGTTCCCTCTAATAGCTTTGGATGGGTCA	3600
CO	CAAGATCATTCCAACCGATGTCAAGATCCAGGTGCCACCCCAATTCATTGGTTGGGTGA	3600
WT	CTGGGAAATCATCAATGGCAAAACAGGGTTTATTAATTAATGGAGGAATTAATTGATGAAG	3660
CO	CCGGAAAGTCCAGCATGGCTAAGCAGGGTCTTCTGATTAAAGGGGAATCATTTGATGAAG	3660
WT	GATATACAGGAGAAATACAAGTGATATGTACTAATATTGGAAAAAGTAAATTTAAATTAA	3720
CO	GATACACCGGCAAAATCCAGGTGATCTGCACAAATATCGGCAAAAGCAATTTAAGCTTA	3720
WT	TAGAGGGACAAAATTTGCACAATTAATTTACTACAGCATCCTCAAATTCAGACAGC	3780
CO	TGSAAGGGCAGAGTTGCTCAACTCATCACTCCACACCAAGCAGCAATTCAGACAGC	3780
WT	CTTGGGATGAAAATAAATATCTCAGAGAGGGGATAAAGGATTTGGAAGTACAGGAGTAT	3840
CO	CTTGGGACGAAAACAAGATTAGCCAGAGAGGTGACAAAGGGCTTCGGCAGCAGAGTGTGT	3840
WT	TCTGGGTAGAAAATATTACAGGAGCACAAAGTGAACATGAGAAATTGGCATACATCAACAA	3900
CO	TCTGGGTGAGAAACATCCAGGAGCACAGGACGAGCACGAGAAATTGGCACACCTCCCTTA	3900
WT	AGATATTGGCAAGAAATTATAGATACCAATTGACTGTAGCAAAACAGATAACTCAAGAAT	3960
CO	AGATTTTGGCCCGCAATTACAAGATCCCACTGACTGTGGCTAAGCAGATCACACAGGAAT	3960
WT	GTCTCATTTGCACTAAGCAAGGATCAGGACCTGCAGGTTGTGTCTAGAGATCTCCTAATC	4020
CO	GCCCCCATGCAACCAACAGGTTCTGGCCCCCGGCTGGTGTAGAGTCCCCCAATC	4020

WT	ATTGGCAGGCAGATTGCAACAATTTGGACAAATAGATAATATTGACTTTGTAGAGTCAA	4080
CO	ACTGGCAGGCAGATTGCAACCACTCGACAAACAAATTATCCTGACCTCGTGGAGAGCA	4080
WT	ATTCTAGGATACATACATGCTACATTATTGTCTAAAGAAAATGCAATTATGTACTTCATTGG	4140
CO	ATTCTGGCTACATCCACGCAACACTCTCTCTCCAGGAAAATGCAATTGTGCACCTCCCTCG	4140
WT	CTATTTTGAAGTGGGCAAGATTCTTTTCAACCAAGTCCCTTACACACAGATAACGSCACTA	4200
CO	CAATTCTTGAATGGGCGAGGCTGTTCTCTCCAAATCCCTGCACACCGACAAACGCGCA	4200
WT	ATTTTGTGGCAGAACCAAGTTGTAAATTTGTGAAATTCCTAAAGATAGCACAATCCACAG	4260
CO	ACTTTGTGGCTGACCTGTGTGAATCTGCTGAAGTTCCTGAAAATCGCCCAACCACTG	4260
WT	GAATACCATATCATCCAGAAAGTCAGGCTATTGTAGAAAGGGCAAATAGGACCTTGAAAG	4320
CO	GCATTCCTTATCACCTGAAAGCCAGGGCATTTGTGAGAGGGCCAACAGAACTCTGAAAG	4320
WT	AGAAGATTCAAAGTCATAGACACACACTCAAACACTGGAGGCAGCTTTACAACCTGCTC	4380
CO	AAAAGATCCATCTCAACAGACAATACACAGACATTGGAGGCGGCACCTTCAGCTCGCCC	4380
WT	TCATTACTTGTAAACAAGGGAGGGAAGTATGGGAGGACAGACACCATGGGAAGTATTTA	4440
CO	TTATCAGCTGCAACAAGGAAGAGAAAGCATGGGCGGCCAGACCCCTGGGAGGTCTTCA	4440
WT	TCACTAATCAAGCACAAGTAATACATGAGAACTTTTACTACAGCAAGCACAATCCTCCA	4500
CO	TCACTAACCAGGCCCGGCTCATCCATGAAAAGCTGCTCTTGCAGCAGGCCCACTCCTCCA	4500
WT	AAAAATTTTGTTTTACAAAATCCCTGGTGAAATGATTGGAGGGACCTACTAGGGTGC	4560
CO	AAAAGTTCTGCTTTTATAAGATCCCGGTGAGCAGACTGGAAAGTCTACAGAGTTT	4560
WT	TGTGGAAAGGTGATGTTGCACTAGTAGTTTAATGATGAAGGAAGGGAATAATTGCTGTAC	4620
CO	TGTGGAAAGGAGACGCGCACTTGTGTGAACGATGAGGCGAAGGGATCATCGCTGTGC	4620
WT	CATTAAACAGGACTAAGTTACTAATAAAACCAAAATTGA	4658
CO	CCCTGACACGCAACCAAGCTTCTCATCAGGCCAACTGA	4658



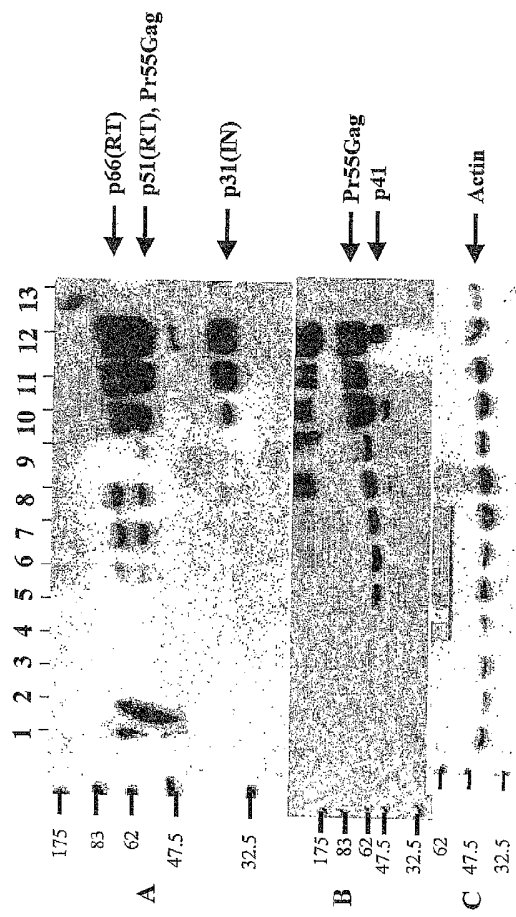


Figure 11

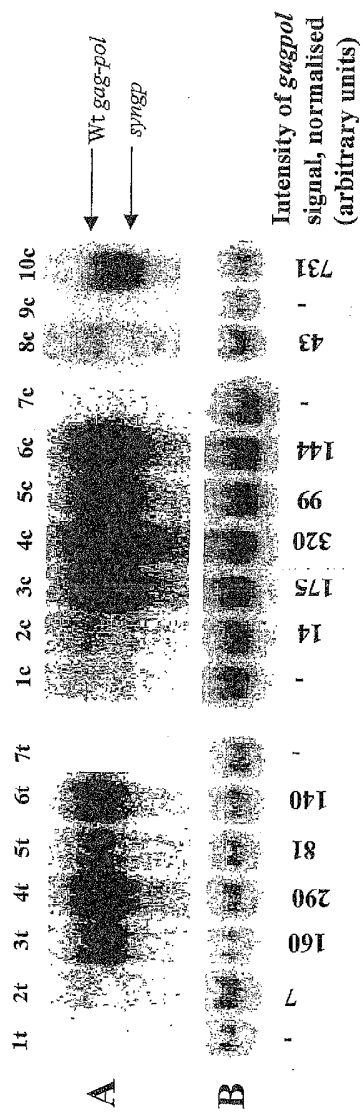


Figure 12

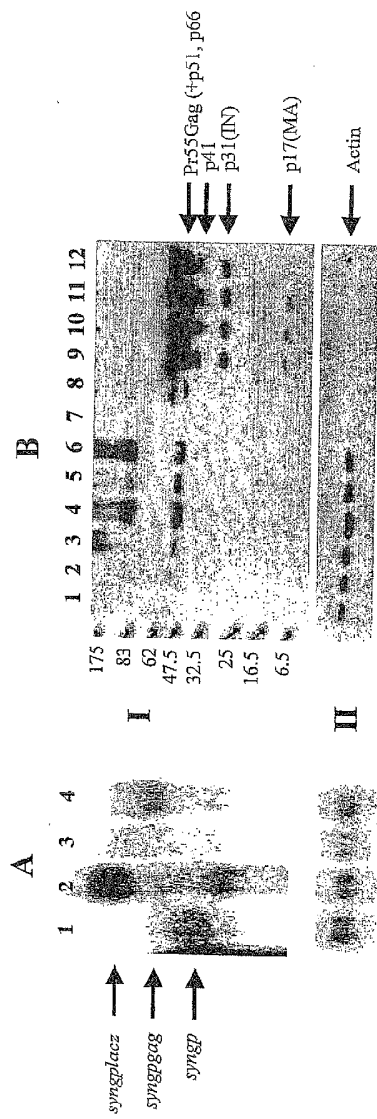


Figure 13

pSYN6 CMVp—Syn *gp*—HXB2 *gag*—pA
 pSYN7 CMVp—Syn *gp*—HXB2 *gag,r*—pA
 pSYN8 CMVp—Syn *gp*—*Lac Z*—pA
 pSYN9 CMVp—HXB2 *gag*—Syn *gp*—pA
 pSYN10 CMVp—HXB2 *gag*—Syn *gp*—RRE—pA
 pSYN11 CMVp—HXB2 *gag,r*—Syn *gp*—RRE—pA
 pSYN12 CMVp—HXB2 *gag*-ΔATG—Syn *gp*—pA
 pSYN13 CMVp—HXB2 *gag*-fr.sh.—Syn *gp*—pA
 pSYN14 CMVp—HXB2 *gag*-ΔATG—Syn *gp*—RRE—pA
 pSYN15 CMVp—HXB2 *gag* 625-1503—Syn *gp*—pA
 pSYN17 CMVp—HXB2 *gag* 1-625—Syn *gp*—pA

Figure 14

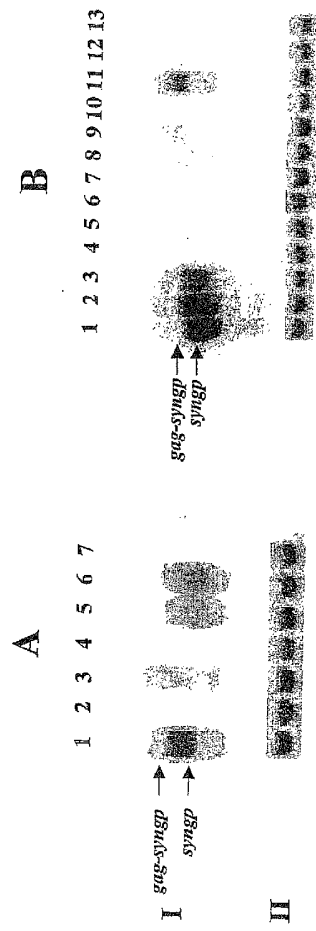


Figure 15

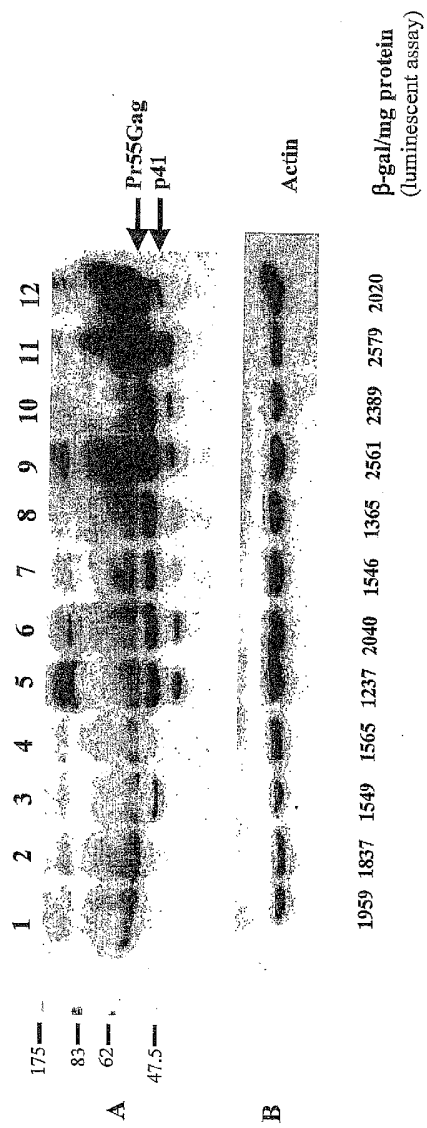


Figure 16

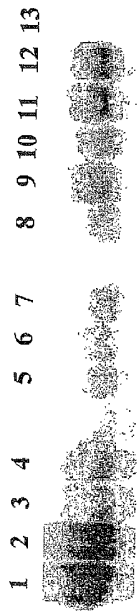
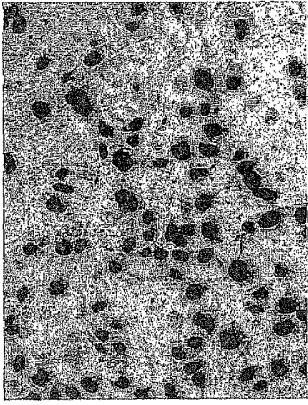
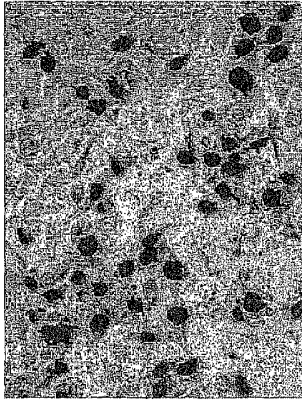


Figure 17

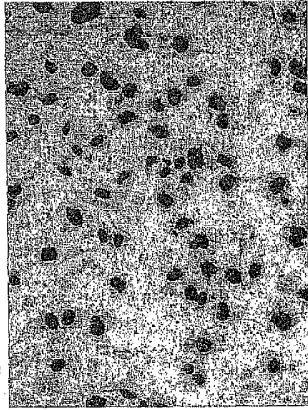
B. pH6nZ (15µg) + pSYNGP (0.5µg)



D. pH53nZ (15µg) + pSYNGP (0.5µg)



A. pH6nZ (15µg) + pSYNGP (5µg)



C. pH6nZ (15µg) + pGP-RRE3 (5µg)

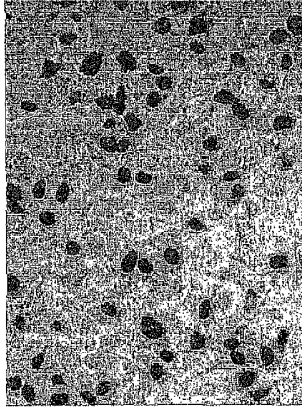


Figure 18

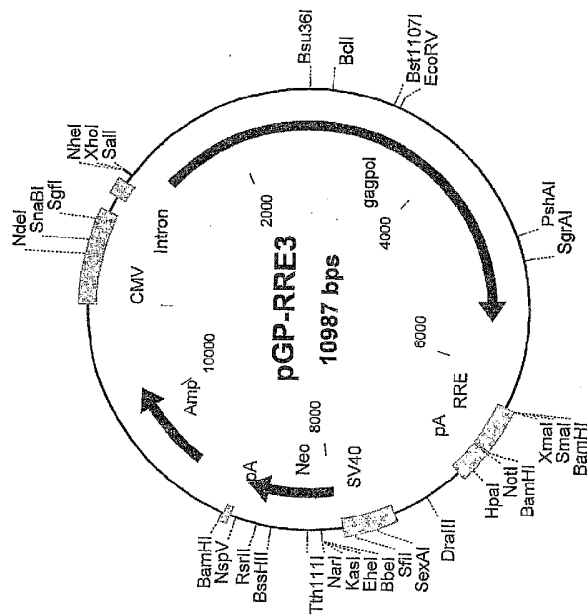


Figure 19

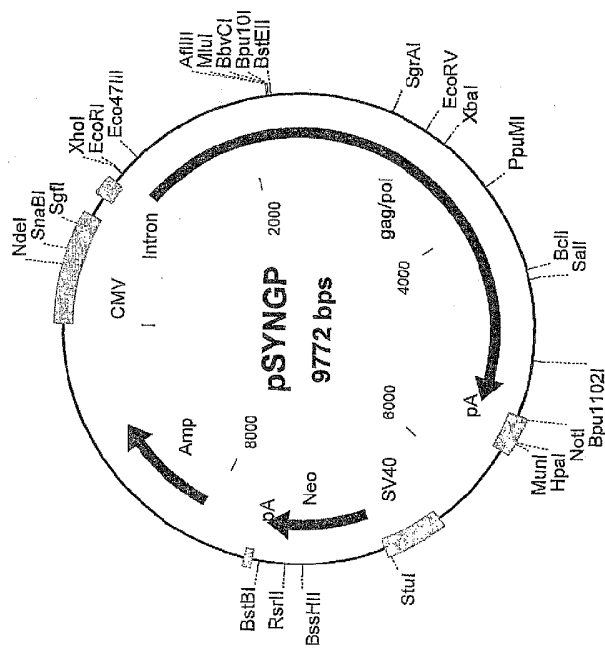


Figure 20

FIGURE 21

Amount of pH6nZ (μ g)	Ratio of <i>gag/gag</i> expression plasmid to pH6nZ (μ g/ μ g)	Titre, pGP-RR3	Titre, pSYNGP	Titre, pSYNGP-RRR	Titre, pSYNGP-ERR
5	1:1	6×10^5	5×10^5	4×10^5	3×10^5
12.5	1:2.5	8×10^5	6.5×10^5		
10	1:10	3×10^5	4×10^5		
5	1:20	3.5×10^5	2.5×10^5		
12.5	1:50	3.5×10^4	3×10^5	3×10^5	3×10^5
18	1:180	2.5×10^4	3×10^5		

Figure 22

Gag-pol expression vector	Vector genome	Rev expression	Titre (IU/ml) ^a
pGP-RRE3	pH6nZ	Yes / from pH6Z	$8.0 \pm 0.2 \times 10^5$
pGP-RRE3	pH6.1nZ	No	$2.1 \pm 0.3 \times 10^3$
pGP-RRE3	pH6.1nZ	Yes / from pCMV-Rev	$4.1 \pm 0.3 \times 10^4$
pGP-RRE3	pH6.2nZ	No	$1.7 \pm 0.4 \times 10^3$
pGP-RRE3	pH6.2nZ	Yes / from pCMV-Rev	$5.6 \pm 0.3 \times 10^4$
PSYNGP	pH6nZ	Yes / from pH6Z	$7.8 \pm 0.2 \times 10^5$
PSYNGP	pH6.1nZ	No	$5.4 \pm 0.3 \times 10^4$
PSYNGP	pH6.1nZ	Yes / from pCMV-Rev	$5.2 \pm 0.4 \times 10^4$
PSYNGP	pH6.2nZ	No	$5.0 \pm 0.2 \times 10^4$
PSYNGP	pH6.2nZ	Yes / from pCMV-Rev	$5.5 \pm 0.3 \times 10^4$

Figure 23

Vector genome	Gag length (nt)	SD mutation	PSYNGP	pGP-RRE3	pGP-RRE + Rev *
pH6nZ	625	No	$6.7 \pm 0.7 \times 10^5$	Not done	Not done
pH6.1nZ	625	No	$3.2 \pm 0.1 \times 10^4$	Not done	Not done
pH6.2nZ	527	No	$2.7 \pm 0.3 \times 10^4$	Not done	Not done
pHS1nZ	40	No	$2.1 \pm 0.4 \times 10^4$	$2.1 \pm 0.9 \times 10^3$	$1.4 \pm 0.3 \times 10^4$
pHS2nZ	260	No	$1.5 \pm 0.3 \times 10^4$	$4.2 \pm 0.6 \times 10^3$	$3.0 \pm 0.5 \times 10^4$
pHS3nZ	360	No	$1.3 \pm 0.3 \times 10^5$	$6.5 \pm 0.7 \times 10^3$	$1.7 \pm 0.4 \times 10^5$
pHS4nZ	625	No	$1.0 \pm 0.9 \times 10^4$	$1.2 \pm 0.7 \times 10^3$	$1.0 \pm 0.3 \times 10^4$
pHS5nZ	40	Yes	$2.0 \pm 0.5 \times 10^4$	$6.0 \pm 0.8 \times 10^2$	$2.3 \pm 0.5 \times 10^4$
pHS6nZ	260	Yes	$1.0 \pm 0.2 \times 10^4$	$8.1 \pm 0.6 \times 10^2$	$1.0 \pm 0.4 \times 10^4$
pHS7nZ	360	Yes	$1.9 \pm 1 \times 10^4$	$2.4 \pm 0.9 \times 10^3$	$4.2 \pm 0.8 \times 10^4$
pHS8nZ	625	Yes	$8.0 \pm 1.0 \times 10^3$	$1.0 \pm 0.5 \times 10^3$	$4.0 \pm 0.9 \times 10^3$

Figure 24

Vector genome	PSYNGP	PSYNGP + Rev *
pH6nZ	6.5×10^5	6.9×10^5
pH6.1nZ	1.5×10^4	1.7×10^4
pH6.1nZR	2.3×10^3	1.6×10^5
pHS1nZ	8.5×10^3	8.4×10^3
pHS1nZR	8.1×10^3	8.8×10^4
pHS3nZ	1.2×10^5	1.4×10^5
pHS3nZR	2.5×10^3	4.8×10^5
pHS7nZ	7.3×10^3	7.0×10^3
pHS7nZR	3.4×10^3	5.8×10^3

Figure 25

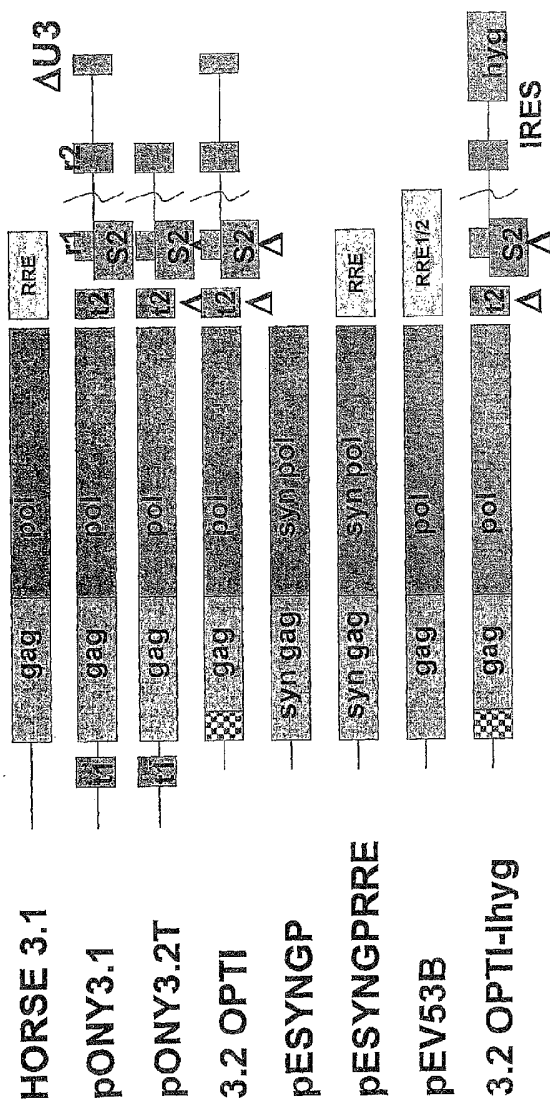
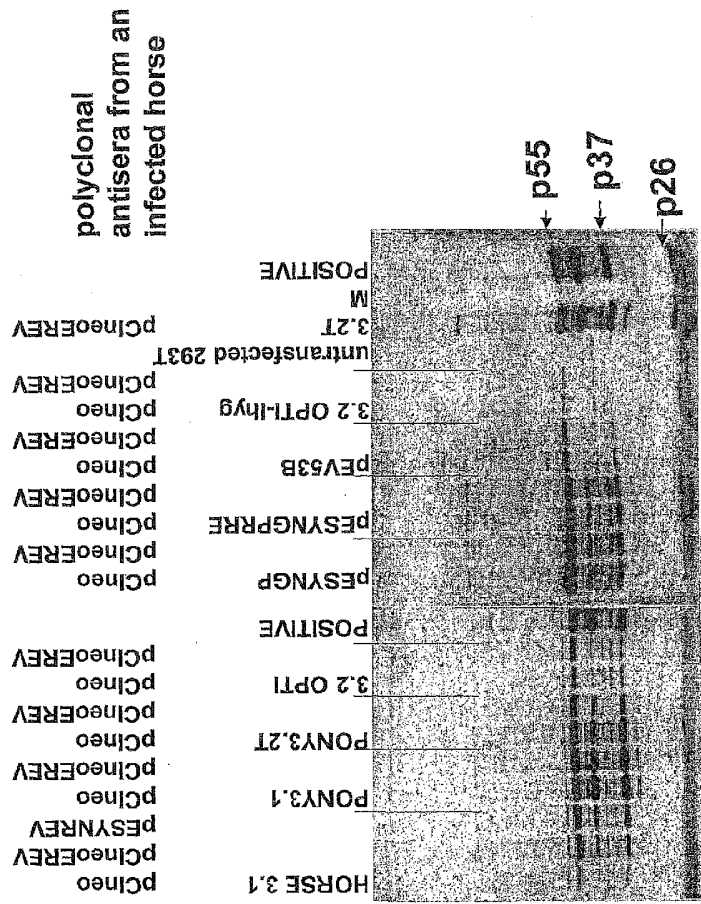


Figure 26
Western blot of 293T extracts



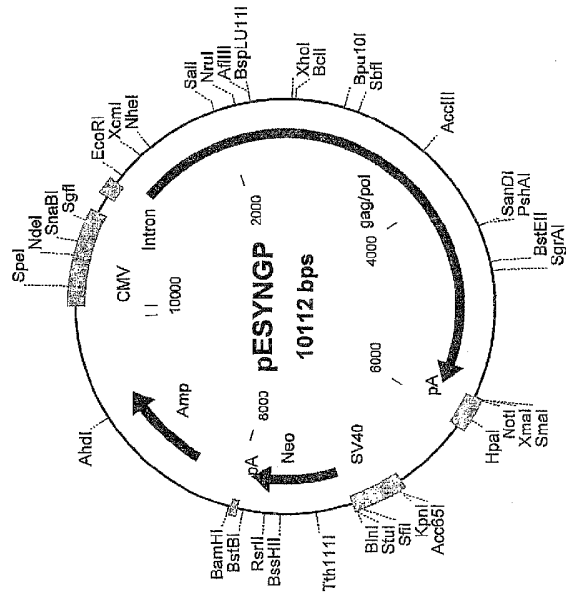


Figure 27

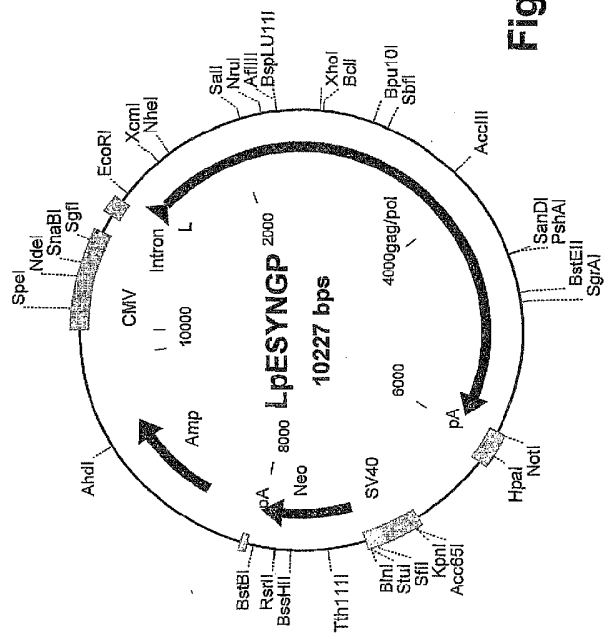


Figure 28

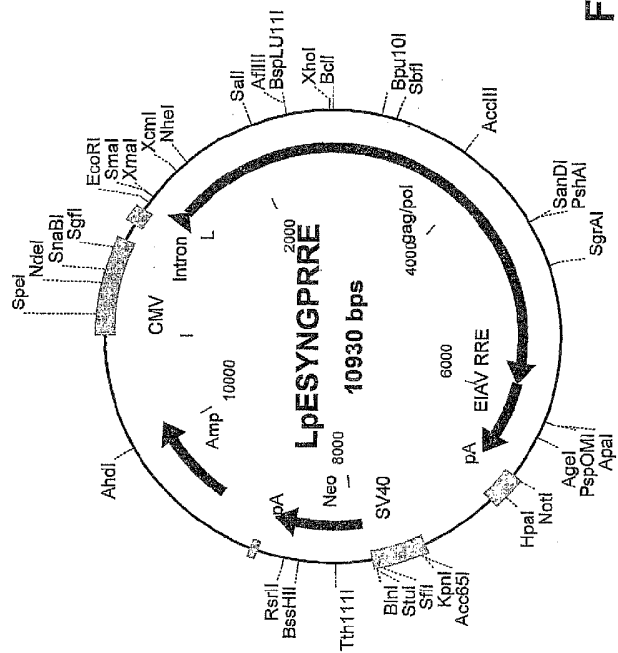


Figure 29

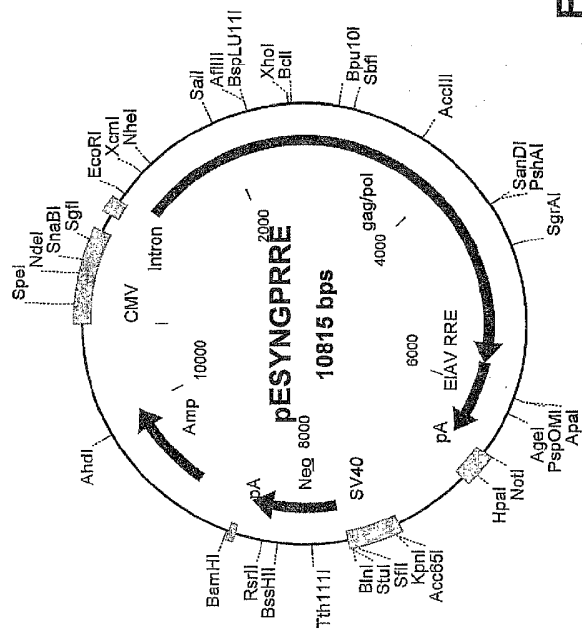


Figure 30

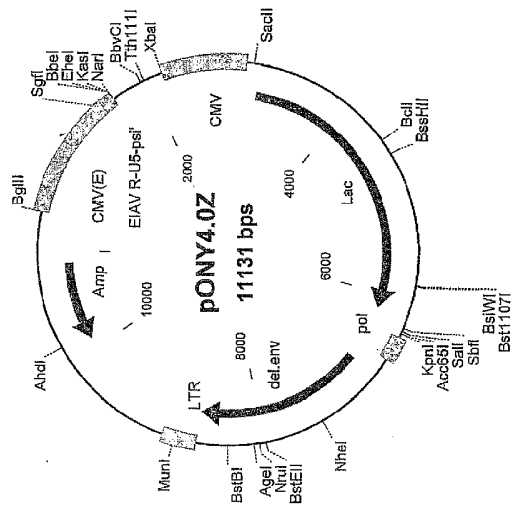


Figure 31

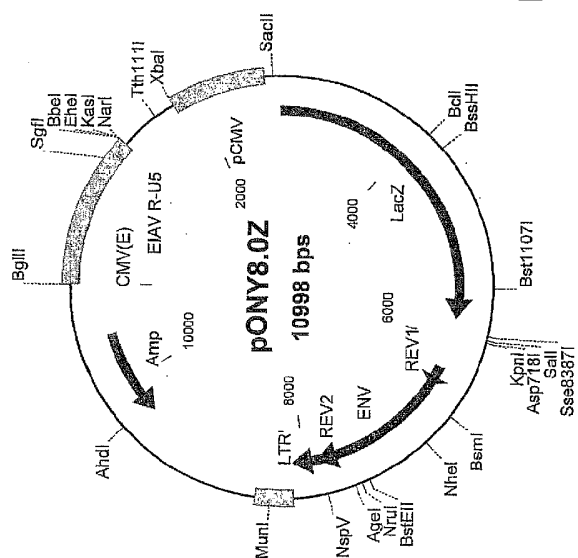


Figure 32

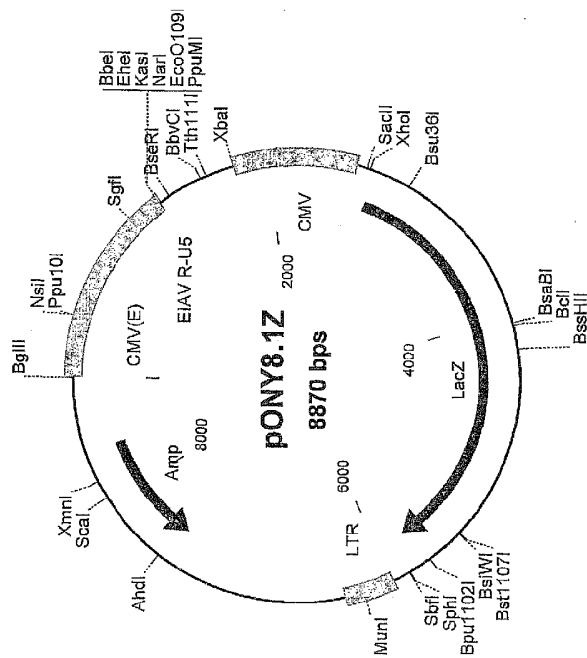


Figure 33

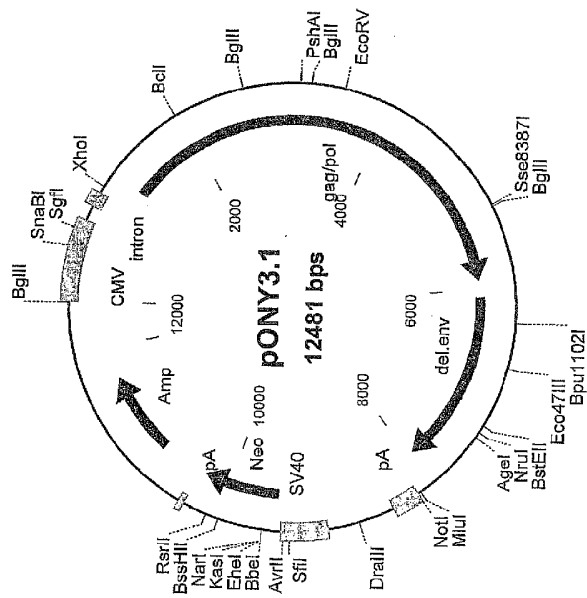


Figure 34

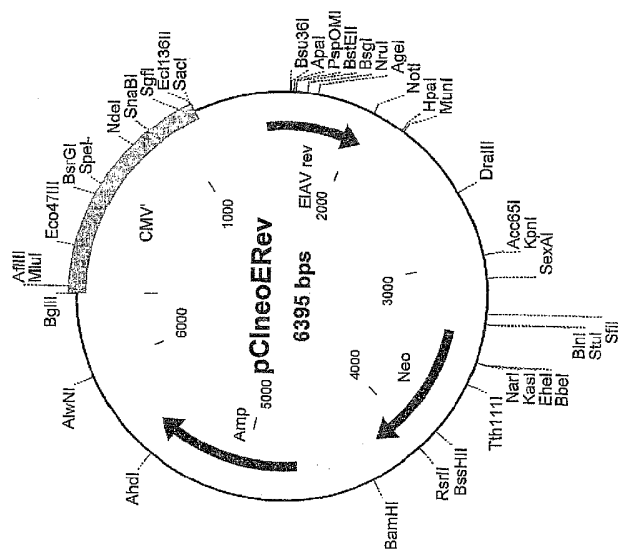


Figure 35

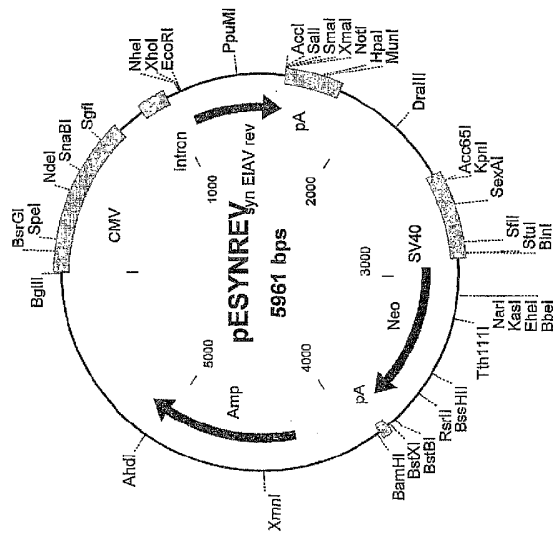


Figure 36

	Genome	Other	TU/ml 1	TU/ml 2	Mean
FIGURE 37					
GAG-POL					
pONY3.1	pONY4.0Z	none	1.84E+06	2.50E+06	2.01E+06
pESYNGP	pONY4.0Z	none	9.00E+05	1.04E+06	9.67E+05
pHORSE3.1	pONY4.0Z	pCIneo	8.80E+04	1.26E+05	1.07E+05
pHORSE3.1	pONY4.0Z	pCIneoREV	1.48E+05	1.42E+05	1.45E+05
pONY3.1	pONY8.0Z	none	1.66E+06	1.48E+06	1.61E+06
pESYNGP	pONY8.0Z	none	2.06E+05	1.10E+05	1.91E+05
pESYNGP	pONY8.0Z	pCIneo	4.00E+04	3.40E+04	2.93E+04
pESYNGP	pONY8.0Z	pCIneoREV	6.40E+04	8.20E+04	6.47E+04
pHORSE3.1	pONY8.0Z	pCIneo	2.00E+03	2.00E+03	2.00E+03
pHORSE3.1	pONY8.0Z	pCIneoREV	1.80E+05	2.48E+05	2.14E+05

Figure 38



FIGURE 39

GAG/POL	OTHER	RT Activity
pONY3.1	pCIneo	87557.9
pONY3.1	pCIneoREV	147061.5
pESYNGP	pCIneo	45701.3
pESYNGP	pCIneoREV	63610.2
L pESYNGP	pCIneo	91729.9
L pESYNGP	pCIneoREV	88487.9
L pESYNGPRRE	pCIneo	71873.4
L pESYNGPRRE	pCIneoREV	101010.7
pESYNGPRRE	pCIneo	32823.1
pESYNGPRRE	pCIneoREV	44863.4
Untransfected 293T cells	NA	1024.6
No cells	NA	1330.6

FIGURE 40

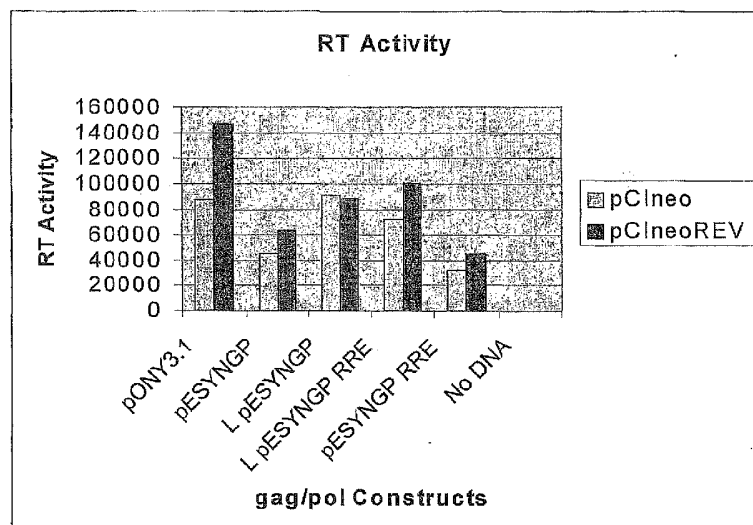


FIGURE 41

VECTOR	GAG/POL	OTHER	titre/ml 1	titre/ml 2	mean
pONY8.0Z	pONY3.1	pCIneo	3.20E+05	2.68E+05	2.94E+05
pONY8.0Z	pONY3.1	pCIneoREV	3.00E+05	5.00E+05	4.00E+05
pONY8.0Z	pESYNGP	pCIneo	4.00E+04	5.60E+04	4.80E+04
pONY8.0Z	pESYNGP	pCIneoREV	2.22E+05	2.40E+05	2.31E+05
pONY8.0Z	L pESYNGP	pCIneo	6.00E+04	6.60E+04	6.30E+04
pONY8.0Z	L pESYNGP	pCIneoREV	2.80E+05	2.70E+05	2.75E+05
pONY8.0Z	L pESYNGPRRE	pCIneo	9.20E+04	6.60E+04	7.90E+04
pONY8.0Z	L pESYNGPRRE	pCIneoREV	3.68E+05	2.82E+05	3.25E+05
pONY8.0Z	pESYNGPRRE	pCIneo	3.20E+04	1.40E+04	2.30E+04
pONY8.0Z	pESYNGPRRE	pCIneoREV	1.20E+05	1.40E+05	1.30E+05

FIGURE 42

GAG/POL	VECTOR	OTHER	titre/ml 1
pONY3.1	pONY8.0Z	pCIneo	1.2E+06
pONY3.1	pONY8.0Z	pCIneoERev	1.0E+06
pONY3.1	pONY8.1Z	pCIneo	1.0E+05
pONY3.1	pONY8.1Z	pCIneoERev	8.0E+05
pESYNGP	pONY8.0Z	pCIneo	2.3E+05
pESYNGP	pONY8.0Z	pCIneoRev	1.0E+05
pESYNGP	pONY8.1Z	pCIneo	1.0E+04
pESYNGP	pONY8.1Z	pCIneoRev	5.1E+04

Figure 43

Comparison of PONY3.1 (nt1246-1606) (TOP) and PONY3.2opt1 (nt366-726) (BOTTOM)

```

atgggagacccctttgacatggagcaaggcctcaagaattagagaaagttacaggtacaa
....c..t..cc.c..c..tc...a..c..g..a..ac.g..a..a..c..c..t..g
gggttcagaaaattaaactactgttaactgttaattggcgctaaagtacagtagacttatt
..tagc..a..gc..t..c..a..c..t..c..c.....at.gtec..g..g..tc..t..c
catgataccaactttgtaaaagaagaaggactggcagctgagggatgtcattccattgtctg
..c..c..t..t..c..t..g..g..a..t....a..c..a..c..g..c..cc.ct..
gaagatgtaactcagacgcctgtcaggacaagaagagagagccctttgaagaacatggtgg
..g..c..g..c..a..at....t..g..g..gc..c..a..t..gc..c.....
gcaatttctgtgttaagatggcctccagatttaataatgtagtagtaggaaagcatca
..c..cagc..a..c..a.....g..g..a..c..c..c..g..t..c..t..a..tagc
ttccagctcctaagagcgaaatatgaaaagaagcctgctaataaaaaagcagctcgagccc
..t..a..g..cc.c..t..g..c..g..a..a..c..c..c..g..a..a..c..a..t
tctgaagaatatccaatcattgatag
agc..g..g..c.....

```

Figure 44

pIRES1hyg ESYNGP

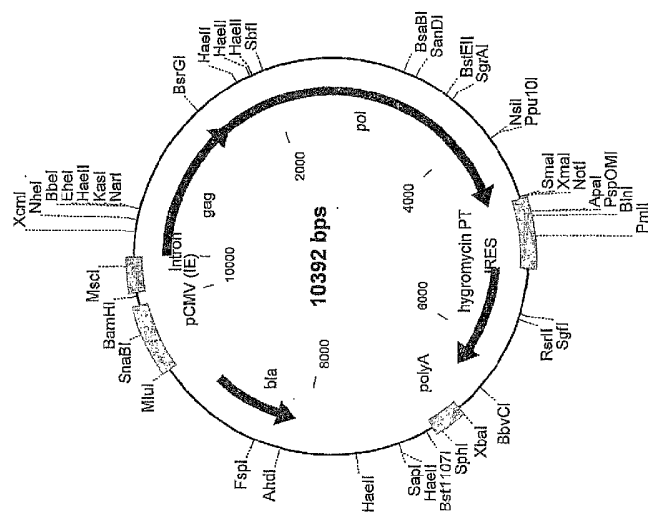


Figure 45

cell line	Ct value	$Ct_{\text{cell line}} - Ct_{\text{pONY8G transient}}$
Q3.29 (3 x 10 ⁶ cells)	17.4	-2.9
8Z.20(24h after induction)	14.35	-6.08
8Z.20(48h after induction)	14.81	-5.62
8Z.20(72h after induction)	16.19	-4.24
pONY8G	20.43	-

Figure 46

transfection number	mass of plasmid transfected (μ g)/10cm diameter dish						titre on D17 cells (LacZ forming units/ml)			
	PESYNGP	PON8.3G	PON8.0Z	PESYNREV	PRV67	pCIneo	1 st harvest	2 nd harvest		
1	16 μ g	16			8		1.4×10^6	1.8×10^5		
2		16			8	16	1.4×10^6	1.6×10^5		
3		16			8		1.4×10^6	2.6×10^5		
4		32			8		2.8×10^6	1.1×10^6		
5			16	8	8		2.4×10^6	2.8×10^5		
6			32	8	8		4.5×10^6	1.1×10^6		

Figure 47

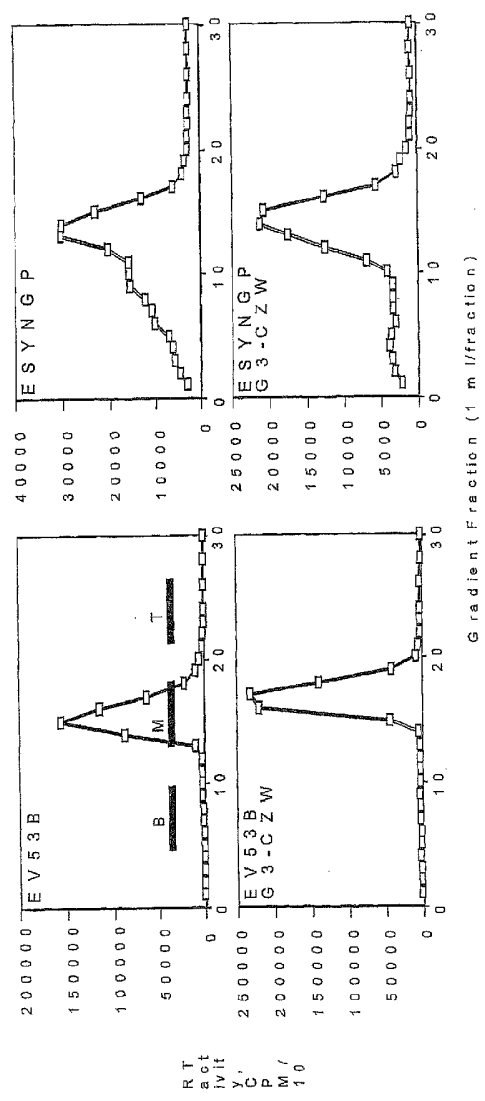


Figure 48

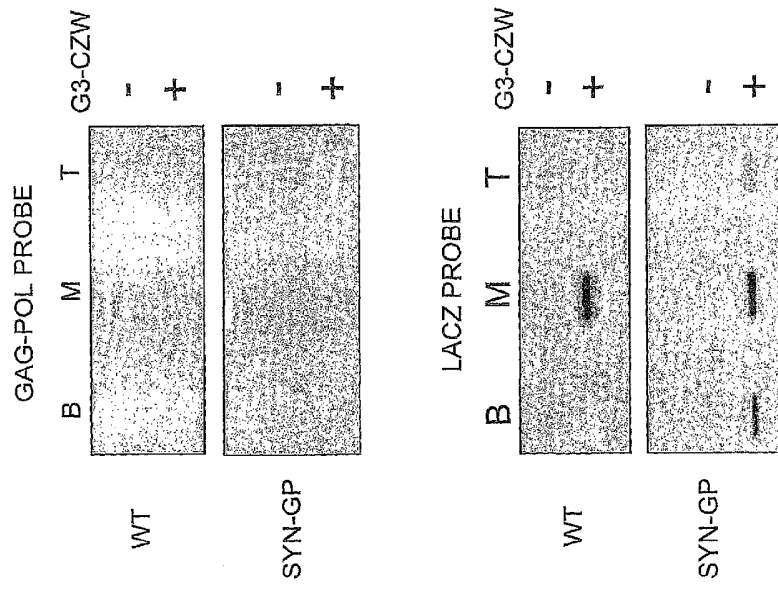


Figure 49

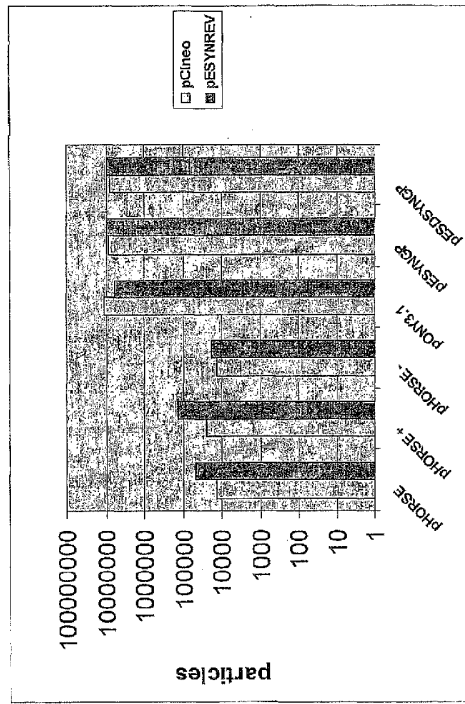


FIGURE 51

